

UNIVERSITY OF HAWAII  
LIBRARY

DEC 4 '57

# ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT  
of the  
PHILOSOPHICAL MAGAZINE

EDITOR

PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD

SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.

PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

VOLUME 6

OCTOBER 1957

NUMBER 24

PRICE per part £1

PRICE per annum £3 15s. 0d. post free

PRINTED AND PUBLISHED BY TAYLOR & FRANCIS LTD  
RED LION COURT, FLEET ST., LONDON E.C.4

QC1  
A36



# *Physics in Medicine and Biology*

The Journal of the Hospital Physicists' Association  
published in association with the Philosophical Magazine

*Editor:*

J. E. ROBERTS, D.Sc.

*Consultant Editor:*

Professor N. F. MOTT, F.R.S.

*Editorial Board:*

W. A. LANGMEAD, M.Sc.

J. S. MITCHELL, C.B.E., F.R.S.

J. ROTBLAT, D.Sc.

D. A. McDONALD, D.M.

G. J. NEARY, Ph.D.

F. W. SPIERS, D.Sc.

J. F. TAIT, Ph.D.

---

## *Contents of July, 1957*

Biophysics Conference (Columbus, Ohio, 1957). By A. Nightingale, M.A., Ph.D., Guy's Hospital Medical School, London

Common Errors in Obtaining and Evaluating Experimental Data. By Maurice H. Belz, Department of Statistics, University of Melbourne

The Dielectric Method of Estimating Protein Hydration. By E. H. Grant, B.Sc., Ph.D., Department of Physics Applied to Medicine, The Middlesex Hospital Medical School, London, W.1

Radio-Caesium in Dried Milk. By D. V. Booker, M.A., Health Physics Division, A.E.R.E., Harwell

The Theory of Tracer Experiments with  $^{131}\text{I}$ -Labelled Plasma Proteins. By Christine M. E. Matthews, Ph.D., B.A., B.Sc., The National Institute for Medical Research, The Ridgeway, Mill Hill, London, N.W.7

Photoelectric and Cinematographic Measurements of the Motility of Bull Sperm Cells. By R. Rikmenspoel and G. van Herpen, Research Institute for Animal Husbandry T.N.O., Utrecht, Netherlands

A Whole-Body Radioactivity Counter for Bed Patients. By J. F. Fowler, Ph.D., M.Sc. Physics Department, King's College Hospital, London, S.E.5

---

Subscription price per volume £3 10s. post free, payable in advance

4 parts per volume—£1 per part

*Printed and Published by*

**TAYLOR & FRANCIS LTD**

**RED LION COURT, FLEET STREET, LONDON, E.C.4**

Orders originating in U.S.A. and Canada should be sent to the  
Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.



# *Journal of Electronics and Control*

A Philosophical Magazine Associated Journal

*Editor:*

J. THOMSON, M.A., D.Sc., M.I.E.E., F.Inst.P.

*Consultant Editor:*

Professor N. F. MOTT, F.R.S.

*Editorial Board:*

Professor P. AIGRAIN (France)  
Professor H. B. G. CASIMIR (Holland)  
Dr. W. KLEIN (Germany)  
Dr. R. KOMPNER (U.S.A.)  
Mr. J. F. COALES (U.K.)

---

## *Contents of September, 1957*

### Electronics Section

Interdigital and Other Slow Wave Structures. By J. C. Walling, Mullard Research Laboratories, Salfords, Surrey

Geometrical Effects in Transverse Magnetoresistance Measurements. By J. R. Drabble and R. Wolfe, Communication from the Staff of the Research Laboratories of the General Electric Company Limited, Wembley, England

An Experimental High Power Pulsed Travelling Wave Tube. By J. F. Gittins, N. H. Rock and A. B. J. Sullivan, Services Electronics Research Laboratory, Admiralty, Baldock, Herts.

Electron Beam Focusing in Three-Anode Guns for Travelling Wave Tubes. By T. S. Chen and L. Kovach, Tube Division, Radio Corporation of America, Harrison, New Jersey

Preferential Diffusion of Sb along Small Angle Boundaries in Ge and the Dependence of this Effect on the Direction of the Dislocation-lines in the Boundary. By F. Karstensen, Zentrallaboratorium, Laboratorium Hamburg, Allgemeine Deutsche Philips Industrie G.m.b.H.

### Book Notices

### Control Section

The Stabilization of Remote Position Control Systems by Proportional Coulomb Damping. By E. A. Freeman, King's College, University of Durham, Newcastle-upon-Tyne

Heating Effects in Single Crystal Barium Titanate. By D. S. Campbell, The Plessey Co. Ltd., Caswell, Towcester, Northants.

---

Price per part £1 5s.

Price per volume £7 post free, payable in advance

6 monthly issues per volume

*Printed and Published by*

**TAYLOR & FRANCIS LTD**

**RED LION COURT, FLEET STREET, LONDON, E.C.4**

Orders originating in U.S.A. and Canada should be sent to the  
Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.



# *The Philosophical Magazine*

First Published in 1798

*Editor :*

PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

*Editorial Board :*

SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.

PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

---

## *Contents of October, 1957*

Liquid-Solid Equilibrium in Copper and Silver Alloys ; with an Appendix on the Equilibrium Diagram of the System Copper-Arsenic. By William Hume-Rothery and Jennifer Burns, Department of Metallurgy, Oxford

The Centre of a Dislocation : I. By Elizabeth H. Yoffe, Cavendish Laboratory, University of Cambridge

Photodisintegration of the Alpha Particle. By B. H. Bransden and A. C. Douglas, Department of Natural Philosophy, University of Glasgow and H. H. Robertson, CERN, European Organization for Nuclear Research, Geneva

Nuclear Forces and the Structure of the Lithium Isotopes. By J. M. Soper, Cavendish Laboratory, Cambridge

Electrostatic Fields round Localized Defects in Metals. By P. M. Lee and N. H. March, Department of Physics, The University, Sheffield

On the Mechanism of Work Hardening in Face-Centred Cubic Metals with Special Reference to Polycrystalline Copper : II. By P. Feltham and J. D. Meakin, The University, Leeds

The Atomic Heats of Titanium, Zirconium and Hafnium. By N. M. Wolcott, Clarendon Laboratory, Oxford, England

The Angular Correlation of the Gamma Cascade in Sulphur 34. By P. S. Fisher and G. G. Shute, Department of Physics, University of Birmingham

Helical Dislocations and Spiral Etch-Pits. By S. Amelinckx, W. Bontinck and W. Dekeyser, Laboratorium voor Kristalkunde, Rozier, 6, Gent, Belgium

The Hardening and Embrittlement of Steel by Irradiation with Neutrons. By A. T. Churchman, Associated Electrical Industries Limited, Aldermaston and I. Mogford and A. H. Cottrell, Atomic Energy Research Establishment, Harwell

On the Electronic Conductivity of Crystals of Silver Halides. By J. W. Mitchell, H. H. Wills Physical Laboratory, University of Bristol

The Thermal Instability of a Rotating Fluid Sphere Heated Within : II. By S. Chandrasekhar, F.R.S., University of Chicago

Correspondence :

On Spiral Etch Pits in Germanium and Silicon. By S. G. Ellis, RCA Laboratories, Princeton, New Jersey

Reviews of Books

---

Price per part £1 5s.

Price per annum £13 10s. post free, payable in advance

*Printed and Published by*

**TAYLOR & FRANCIS LTD**

**RED LION COURT, FLEET STREET, LONDON, E.C.4**

# *Journal of Fluid Mechanics*

*Editor:*

Dr. G. K. BATCHELOR, F.R.S.  
Cavendish Laboratory, University of Cambridge, Cambridge, England

*Assistant Editors:*

Dr. T. B. BENJAMIN, Dr. I. PROUDMAN

*Associate Editors:*

Professor G. F. CARRIER  
Pierce Hall, Harvard University, Cambridge 38, Massachusetts, U.S.A.

Dr. W. C. GRIFFITH  
Research and Development Laboratory, Missile Systems Division, Lockheed Aircraft Corporation, Palo Alto, California, U.S.A.

Professor M. J. LIGHTHILL  
Department of Mathematics, The University, Manchester, England

---

## *Contents of August, 1957*

- The Production of Uniform Shear Flow in a Wind Tunnel. By P. R. Owen and H. K. Zienkiewicz, Mechanics of Fluids Department, University of Manchester
- An Exact Solution for Progressive Capillary Waves of Arbitrary Amplitude. By G. D. Crapper, Department of Mathematics, University of Manchester
- The Flow Past a Rapidly Rotating Circular Cylinder in a Uniform Stream. By D. W. Moore, Jesus College, Cambridge
- Experiments on the Onset of Wave Formation on a Film of Water Flowing Down a Vertical Plane. By A. M. Binnie, Trinity College, Cambridge
- Wave Formation in Laminar Flow Down an Inclined Plane. By T. Brooke Benjamin, Department of Engineering, University of Cambridge
- A Note on the Application of the Supersonic Area Rule to the Determination of the Wave Drag of Rectangular Wings. By R. C. Lock, Aerodynamics Division, National Physical Laboratory, Teddington
- Experiments on Convection of Isolated Masses of Buoyant Fluid. By R. S. Scorer, Department of Meteorology, Imperial College, London
- The Vorticity Jump Across a Gasdynamic Discontinuity. By Wallace D. Hayes, Department of Aeronautical Engineering, Princeton University, and the Ramo-Wooldrige Corporation
- The Collision of Two Ionized Streams. By F. D. Kahn, Department of Astronomy, Physical Laboratories, University of Manchester

---

Price per part £1

Price per annum £5 10s. post free

*Printed and Published by*

**TAYLOR & FRANCIS LTD**

**RED LION COURT, FLEET STREET, LONDON, E.C.4**

Orders originating in U.S.A. and Canada should be sent to the  
Academic Press Inc., 111 Fifth Avenue, New York, 3, N.Y., U.S.A.



## CONTENTS

The Generation of Static Charge. By W. R. HARPER, 67 Burton Court, London, S.W.3 . . . . .	365
A Theory of the Fracture of Metals. By A. N. STROH, Department of Physics, University of Sheffield . . . . .	418





# ADVANCES IN PHYSICS

## A QUARTERLY SUPPLEMENT

of the

## PHILOSOPHICAL MAGAZINE

---

---

VOLUME 6

OCTOBER 1957

NUMBER 24

---

---

### The Generation of Static Charge

By W. R. HARPER

67 Burton Court, London, S.W.3

#### CONTENTS

- § 1. INTRODUCTION.
- § 2. GAS/LIQUID ENTRAINMENT CHARGING.
  - 2.1. Statistical fluctuation charging of droplets.
    - 2.2.1. Formation of large ions by entrainment : discussion.
    - 2.2.2. Formation of large ions by entrainment : conclusions.
  - 2.3. Motes and veils.
- LIQUID/SOLID CHARGING.
  - 3.1. Charging of a liquid by streaming.
  - 3.2. Charging by freezing.
  - 3.3. Thunderstorms : charging by rime formation.
- § 4. MERCURY.
- § 5. SOLID/SOLID CHARGING.
  - 5.1. Ionic transfer between solids.
  - 5.2. Theory of electronic transfer between solids.
  - 5.3. The two classes of insulators.
    - 5.3.1. Electrophobic insulators.
    - 5.3.2. The charging of electrophilic insulators.
  - 5.4. Metals and oxides.
  - 5.5. The significance of friction.
- § 6. CONCLUSIONS, AND A WORKING HYPOTHESIS.
- ACKNOWLEDGMENTS.
- REFERENCES.

#### § 1. INTRODUCTION

THE appearance of electrification on insulating surfaces that have been rubbed, commonly referred to as frictional electrification, is well known in technological processes, where it frequently gives rise to difficulty or hazard. These manifestations give little indication of what is causing them : one is seldom sure whether a particular process will give charging or not, and often uncertain, when it does, what the sign of charging will be. Henry (1953 a), for example, says that it is no surprise to find the warps of acetate rayon that are being woven on two neighbouring looms, apparently identical in all respects, strongly charged, but with opposite sign. ' Frictional ' charging, in fact, is temperamental to an extent that



friction itself is not, and this suggests that, whereas the force of friction arises from the bulk properties of the materials concerned, the electrification depends on true surface properties, which are themselves critically dependent on traces of contamination and so on. The name 'Triboelectrification' is an unfortunate one, since it begs an important question, and probably incorrectly; Volta, in 1779, taught that the virtue of friction was merely to multiply the points of contact, though it can also, of course, alter the surface conditions at them. It would, however, be premature to alter the name to contact electrification, as suggested by Peterson (1954 a).

It is clear, then, that carefully controlled laboratory experiments are needed to identify the physical or chemical processes that are capable of giving rise to charging, but this is not all. In certain experiments of the author's (Harper 1953 a), polystyrene was found to be extremely reluctant to charge, but, in the manufacture of capacitors, polystyrene film charges freely under seemingly similar conditions, and is thereby rendered awkward to handle. Another kind of experiment is required to find out what is happening in the factory that did not happen in the laboratory. Indeed, there are three aspects of the generation of static charge requiring separate investigation. Firstly, there is the collection of empirical facts from a wide range of technological experience (for a review see Arthur 1955). Secondly, there are laboratory experiments carried out under carefully controlled conditions. Thirdly, there is the experimental demonstration of which given physical or chemical processes actually operate in what circumstances. Published work too often fails to conform to these requirements, new facts and ideas being produced that fail to take their place in any scheme, though, of course, the requirements are very difficult to fulfil. Recent work seems, at first sight, merely to add complexity to an already confused subject, and a feeling has grown up (Henry 1953 b, and a review by Loeb 1945) that quite a number of processes may combine, differently according to circumstances, in producing static electrification. If this be so, then the explanation sought is a complicated pattern of which we now see only a few fragments. There are, however, threads of evidence discernible in certain papers which lead to a somewhat simpler picture capable of correlating different manifestations of static electrification. Emphasis will therefore be placed on these papers, and the subject ordered round the development of a working hypothesis.

## § 2. GAS/LIQUID ENTRAINMENT CHARGING

When a gas is bubbled through a liquid some of the liquid is entrained in the stream of gas in the form of particles of a wide range of size. The largest drops formed may be returned to the liquid, intermediate size groups will be more or less permanently entrained, and there are, furthermore, submicroscopic particles present which are easily detected by their ability to turn into large ions when the surrounding gas is ionized



by a radioactive source. Unless the liquid has been specially purified, many of the particles are electrically charged at formation. Splashing, spraying and the shattering of drops have an effect similar to bubbling. The origin of the charging was the subject of a controversy at the Symposium on Static Electrification held by the Institute of Physics in London in March 1953, so it is worth reviewing the evidence with special care. We shall begin with the droplets of diameter a few microns.

### 2.1. Statistical Fluctuation Charging of Droplets

If electrolytic ions are present in the liquid, and the entrainment process results in a random sample of bulk liquid being transferred direct to the gas as a droplet, then droplets of a given size will contain, on the average, as many positive as negative ions; but, owing to the irregular distribution of ions in the liquid, some droplets will acquire an excess of positive ions, and some of negative. Dodd (1953), following the earlier work of Chapman (1934) and Natanson (1949) have investigated charging of this kind, the mathematical theory having already been worked out by Bateman (1911). The liquids were atomized by spraying, and charges on the droplets investigated using the Millikan oil-drop method, or later modifications of it.

Natanson found detailed agreement between theory and experiment. He points out that the theory is only applicable when the conductivity of the liquid is too low for significant changes in ionic concentration to occur during the intermediate stages of formation of the droplets. This implies badly conducting liquids of low dielectric constant  $D$ . They should contain considerably fewer than†  $2 \times 10^{12} D$  ions/cm<sup>3</sup>, if we are considering droplets of radius about a micron. At such ionic concentrations the charge does not depend on the initial shape of the 'droplet-to-be'. The statistics of the sampling will be disturbed if the droplets change their size before measurement because of evaporation, if there is coagulation in the mist, or because of other disturbances. This aspect of the investigation was carefully considered by Dodd. It appears that all such considerations have probably been adequately met both by Dodd and by Natanson. The latter worked with ions of triethanolamine oleate in specially purified transformer oil; it is uncertain whether the ions in the liquids used by Dodd came from the dissociation of these liquids, or from unknown impurities.

Let there be  $N$  molecules of formula  $A_p B_q$  per cm<sup>3</sup>, dissociated into  $p$  positive ions each carrying a charge  $z_+$  elementary units, and  $q$  negative ions each carrying a charge  $z_-$ , and‡ let  $w = pz_+^2 + qz_-^2$ . For sufficiently many ions in a droplet, the asymptotic formula for the probability of a droplet of volume  $v$  containing a net charge of  $\sigma$  elementary units is

$$(2\pi Nvw)^{-1/2} \exp(-\sigma^2/2Nvw).$$

---

† Natanson's  $7 \times 10^{11} D$  seems to be a mistake.

‡  $w = pz_+^2 + qz_-^2$ , in Natanson's paper, is clearly a mistake.

The mean square charge is  $\overline{\sigma^2} = Nvw$ , and the mean absolute charge  $|\overline{\sigma}| = (2Nvw/\pi)^{1/2}$ . For few ions in a droplet, the mean square charge is the same as the asymptotic value, but the mean absolute charge becomes, for  $w=2$ ,

$$|\overline{\sigma}| = 2 \exp(-2Nv) \sum_{\sigma=0}^{\infty} \sigma I_{\sigma}(2Nv),$$

where  $I_{\sigma}$  is the Bessel function of imaginary argument and order  $\sigma$ .

Natanson found that the distribution of charge among droplets of a given size had a maximum for zero charge, and fell off symmetrically on both sides of the maximum. Dodd concluded that, for di-n-butyl sebacate droplets of radius a few microns, the distribution curve was Gaussian, in conformity with theory, and the mean absolute charge proportional to the three halves power of the radius ( $r^{3/2}$ ) between one and twenty-five microns. Natanson found  $\overline{\sigma^2}/v$  to be roughly proportional to the conductivity of solutions of triethanolamine oleate in transformer oil (of relatively low conductivity), attributing a systematic departure from proportionality to a change in ionic mobility with concentration arising from molecular association. His results are plotted in fig. 1. He also obtained an elegant check on the theory by confirming the departure of the mean absolute charge from the asymptotic value, for small charges. The experimental values of  $\overline{\sigma^2}$  as a function of radius were fitted to a cube law, for four values of concentration, and are shown in fig. 2. The values of  $\overline{\sigma^2}/v = 2N$  (assuming monovalent ions) were then used to calculate the theoretical variation of the mean absolute charge with radius. The theoretical curves were compared with experiment and are shown in fig. 3, from which it is seen that the agreement is excellent. Natanson's work makes it clear that a correlation between charging and dielectric constant, suggested by Chapman and reconsidered by Dodd, is only to be expected in so far as the dielectric constant determines the extent of electrolytic dissociation. What really determines the extent of charging of the intermediate droplets of a slightly conducting liquid is the ionic concentration (strictly  $\frac{1}{2}Nw$ ), and this makes the conductivity much more significant than the dielectric constant.

We have seen that the theory of statistical fluctuation charging during spraying is well established in the limited range of conditions in which it should apply, and may therefore be made the basis of an experimental method of measuring ionic concentration (if the ions are monovalent—or  $\frac{1}{2}Nw$  if they are multivalent).

### 2.2.1. *Formation of large ions by entrainment: discussion*

We proceed to consider entrainment charging outside the range of conditions in which it is due to the random sampling of the statistically fluctuating distribution of electrolytic ions by direct transfer into the gas. We remove the restriction of low ionic concentration, so that the limit of size of drop having an appreciable chance of carrying a charge is

lowered, and we remove the restriction of low volatility, again extending our interest in the direction of smaller drops, large ions and small ions. In removing the restrictions, it must be remembered that these greater

Fig. 1

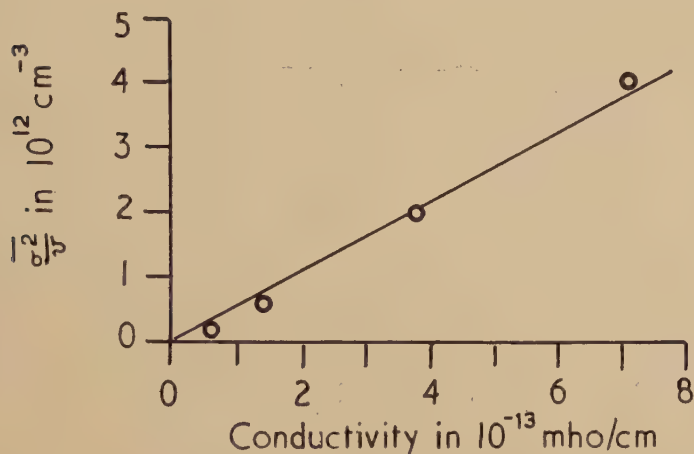
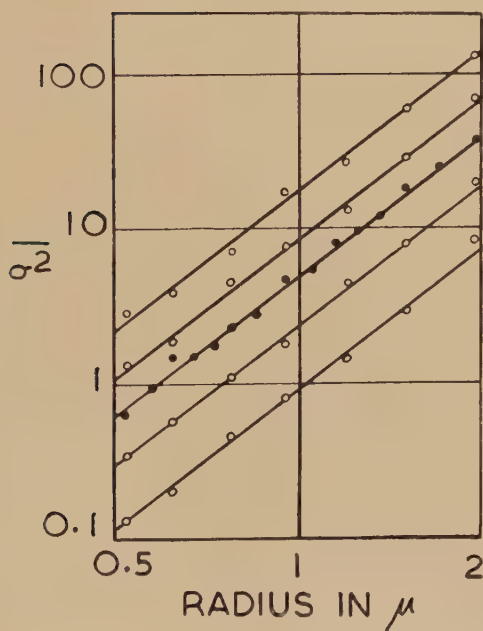


Fig. 2



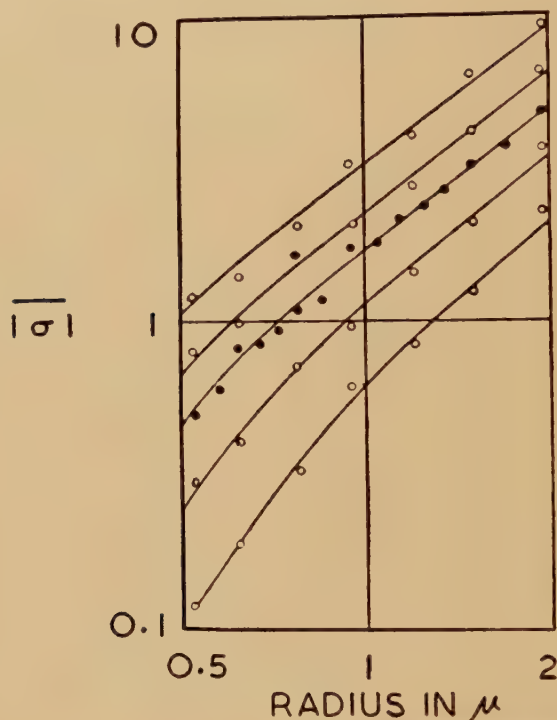
From Zhur. Fiz. Khim.

ionic concentrations can still be found among liquids such as nitrobenzene, normally reckoned as non-conducting, and that nearly all liquids are volatile in the present context. Dodd (1953) used diffusion pump oils to secure low volatility, but, even so, had to apply a correction for the



evaporation of dibutyl phthalate. Water droplets evaporate quickly, as is well known, and Dodd calculated that the change from a diameter of  $100\mu$  to  $1\mu$  occurs in a matter of seconds. The theory of statistical fluctuation charging therefore requires modification for most liquids, even of low conductivity.

Fig. 3



*From Zhur Fiz. Khim.*

The formation of large ions has been known since the earliest experiments: Chapman (1937, 1938) has obtained mobility spectra. Submicroscopic particles cannot be investigated in the same detail as the visible droplets whose size and charge can be individually determined. For the latter, the Millikan method permits a determination of the proportion charged, and by how many elementary units: there is no experimental restriction on charge determination. The amount of charge carried by the submicroscopic particles, however, must be measured in bulk, and a limit is set by the sensitivity of the apparatus. The most sensitive apparatus so far used is that described by Harper (1953 b); even so, for sufficiently non-conducting liquids the charging disappeared, indicating that it had arisen from electrolytic ions already present in the liquid. This was the case for hexane, octane, decane, benzene, xylene, pinene, carbon tetrachloride and a mixture of xylene and carbon tetrachloride, providing they had been purified so as to reduce their conductivity to below  $10^{-13}$  mho  $\text{cm}^{-1}$ . Before purification, pinene

gave plenty of charging, the electrometer drift corresponding to  $+1.9 \times 10^{-4}$  V/sec when collecting positive ions, and  $-1.4 \times 10^{-4}$  V/sec when collecting negative ions, with a gas flow of  $10 \text{ cm}^3/\text{sec}$ , and a bubble size around 3 mm in diameter. The capacitance of the apparatus was about 20 cm, from which it may be calculated that the observed charging could be provided by about 120 pairs of ions per  $\text{cm}^2$  of bubble surface. The conductivity of the pinene was  $1.3 \times 10^{-12} \text{ mho cm}^{-1}$ . If the mobility of the ions had a low value like that for transformer oil and triethanolamine oleate in transformer oil, comparison with Natanson's data gives a concentration of  $6 \times 10^{12}$  pairs of ions per  $\text{cm}^3$ ; it would be one-thousandth of this concentration if the ionic mobility was like that in water. Even the lower concentration gives a density of 120 pairs of ions/ $\text{cm}^2$  of bubble surface for a thickness of 2 Å. There is no problem, then, about where the large ions come from: there are plenty of electrolytic ions already present. This is true even for a liquid which is as good an insulator as the specially purified transformer oil used by Natanson. The mechanical subdivision of the liquid by the formation of a bubble film and its subsequent disintegration can easily supply a sufficient number of ions. By disintegrating into particles of linear dimensions comparable with the thickness of the film, large ions might be formed directly; in practice they are more likely to be formed from much bigger sheets of liquid by evaporation. Many more particles will be formed initially during the disintegration of the film than would be expected to contain ions. Some will evaporate completely, but many do not. As found by De Broglie (1909), they give rise to neutral centres, recognizable because they can be detected as large ions if they have been allowed to pick up small ions by proximity with a radioactive source. In the case of pinene (Harper 1953 b), a hundred times as many neutral centres as large ions were detected. The stability of the neutral centres against evaporation suggests that they come from some form of solid contamination, either dissolved or particulate, and many of the large ions formed by entrainment may be those 'neutral centres' which happen to carry a charge. The charge may have been adsorbed on particulate contamination in the liquid, or merely become associated with the contamination because of happening to be in the same piece of film after disintegration.

We are thus led to inquire how far a modified theory of statistical fluctuation charging will account for the principal features of liquid entrainment charging. The most outstanding feature is that the large ions generally carry a net positive or negative charge, the compensating charge being on the larger drops, though the latter may be overlooked for the lack of a suitably sensitive means of detecting it. Modifying the theory involves three considerations.

(i) Evaporation will result in given charges being associated with smaller drops or ions. It cannot, however, give rise to a net charge that was not produced by the entrainment process unless there is an excess loss of ions of one sign to the walls. This was excluded in the case of pinene (Harper 1953 b) already discussed.

(ii) The charge on a large ion can, in principle, depend on the shape of the 'droplet-to-be' from which it was formed. This may be of great importance if there is an ionic double layer at the surface of the liquid. A sufficiently thin film will then contain an excess of the ion that is preferentially adsorbed to the surface, and droplets formed from it carry a net charge of that sign. The disruption of a double layer of charge on the surface of water was first postulated by Lenard (1892) to explain the electrification produced by waterfalls discovered by Elster and Geitel (1890). He originally supposed the negative charge to have been in the air, and the positive in the water, but in a later paper (1915) realized that both components of the double layer must be just inside the water.

(iii) Ionic conductivity may be sufficient to permit a redistribution of charge during the formation of the droplet. Charging of the droplet at the last moment of its formation by electrostatic induction, after the fashion of Kelvin's water-dropper (1867), may arise from contact potentials inevitably present, or from the charge on neighbouring droplets.

Lenard (1915) accounted for the negative excess charge formed from distilled water as coming from a double layer negative outwards, and appears to have visualized some process capable of peeling the negative layer neatly off the positive, because he accounts for the positive particles formed in a minority as arising from electrostatic induction in the field of the more numerous negative particles. Busse (1925) discussed this explanation, but rejected it, preferring to account for the positive particles as having been formed from parts of the surface that had already been denuded of negative ions. Coehn and Mozer (1914) followed the change in the nature of the charge produced by bubbling as traces of electrolytic impurity were added to conductivity water. At first the resultant negative charge imparted to the gas consisted of negative particles only. As impurity was added to the water, positive particles made their appearance, and increased in numbers until the 'inversion point' was reached. At this point the resultant charge was zero, but the gas still contained a lot of both positive and negative particles in equal numbers. It seems quite impossible to account for such behaviour in terms of electrostatic induction. Were it not for the absence of positive ions using conductivity water, there would be no need for the delicate separation of the two components of a double layer: a clumsier process would supply a mixture of ions with one in excess. Gill (1953 a) has emphasized the inherent improbability of Lenard's hypothesis.

Gill and Alfrey (1949 a, 1952 a) suppose other investigators to have been misled when interpreting their experiments as being in accordance with Lenard's hypothesis, through having overlooked the importance of the charging of drops when they impinge on and rebound from a solid surface at which there may be a contact potential which can give rise to electrostatic induction. They explain their own experimental results along such lines, but in doing so become involved in an inconsistency (Harper 1953 b, Discussion). They also interpret an experiment of

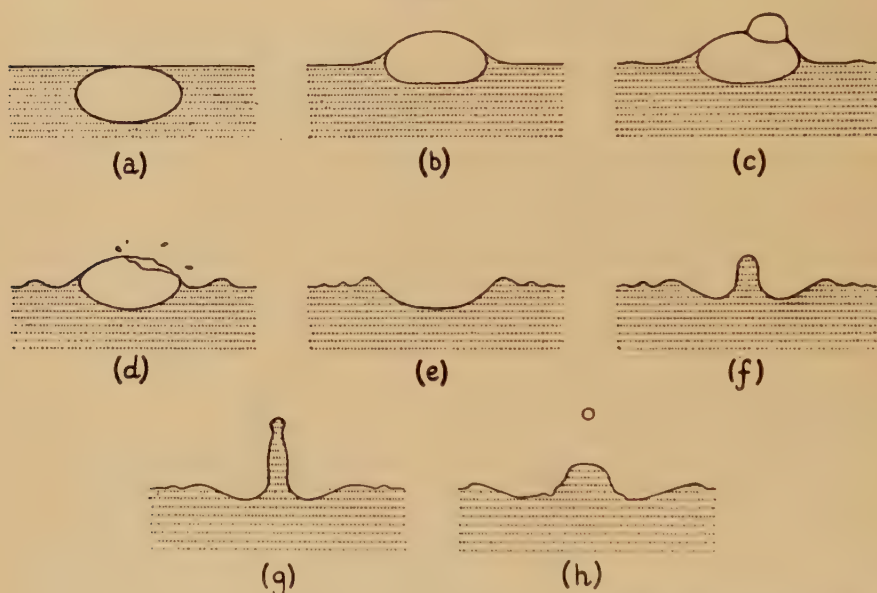


Pomeroy (1908) on spraying in this way, but then make it difficult to understand why the nozzle used in Pomeroy's experiment did not charge up by induction like a water-dropper. Other evidence in support of the orthodox interpretation can be quoted. It may be possible to explain Zeleny's (1933) measurements in the way that Gill and Alfrey suggest, but the earlier work of Simpson (1909), referred to by Zeleny, cannot be explained in this way, as Simpson himself showed by a crucial experiment. This demonstrated that the charge supposed to have been due to the shattering of drops would have been opposite in sign if it had been due to splashing. Bühl (1926), in experiments on the spraying of mercury, investigated the extent to which electrostatic induction affected his results, and concluded that, though present, it played only a subsidiary rôle. The most convincing argument, however, against the charge being produced by induction in the field of a contact potential, was given by Bloch (1911, p. 100). Having observed that a deliberately applied field led to a change in the charging due to bubbling, he then pointed out that the effect observed in the absence of an applied field could not be explained by there being a field already present, because no such field could simultaneously lead to the formation of both positively and negatively charged large ions, and these were observed by him in comparable numbers. The work of Natanson (1949) confirms the conclusion for droplets of intermediate size. We must conclude that electrostatic induction cannot be solely responsible for the charging, though it will certainly lead to complications.

High-speed photography has enabled the mechanics of entrainment processes to be investigated in some detail. We may note, first, that Newitt *et al.* (1954) looked for the rebound of liquid drops from a solid surface, and failed to find it: the drop on impact deformed to a flat disc which subsequently spread to a thin film. This excludes the phenomenon considered important by Gill and Alfrey (1952 a). It should be mentioned that Worthington (1895) investigated the complicated behaviour of a relatively large drop on impact, and even took photographs. It appears from the work of Newitt *et al.* (1954), and that of Woodcock *et al.* (1953), that the bursting of a bubble on the surface of a liquid is a more complicated phenomenon than one might have supposed. Though full agreement on the interpretation of the experiments has not been reached (Blanchard 1954), it is established that two distinct processes of droplet formation are involved: the disintegration of a thin blister that forms on the bubble dome followed by the collapse of the bubble, and, subsequently, the detachment of droplets from a jet of liquid that rises from the crater left after the bubble dome has burst. The sequence of events is shown in fig. 4, and the two critical stages (*c*) and (*h*) in Plate II. Newitt *et al.* find that 3 mm diameter air bubbles bursting at the surface of water produce about one droplet per bubble, of mean diameter 0.7 mm, caused by detachment from the jet. Such large droplets could not escape from a bubbler on the type used by Harper

(1953 b), and may be ignored when looking for the mechanism that generates large ions, though they are the important ones when considering liquid entrainment in industrial plant. According to Blanchard (1954) bubbles of diameter less than  $50\mu$  give rise to jets that detach droplets of diameter between 2 and  $20\mu$ , but these cannot be relevant to our search, because tiny bubbles are extremely difficult to make, and have seldom been used in experiments on entrainment charging. It is to be presumed that the large droplets detached from jets following the bursting of bubbles of medium size will be accompanied by Plateau's spherules, the formation of which is shown in Plate II. These, however, though much smaller than the large droplets, are likely to be in the intermediate range, bigger than large ions. Turning now to the bursting of the blister on the bubble dome, Newitt *et al.* (1954) find that the débris that is ejected when it bursts is recognizable because it is scattered with considerable violence in all directions, instead of being projected vertically upwards.

Fig. 4



From *Trans. Instn. chem. Engrs.*

#### Mechanism of the burst of an air bubble on the surface of water.

The droplets, moreover, give a separate peak in the size distribution curve. Up to ten droplets per bubble burst were observed; many more may have been formed but lost through evaporation. The method of detection (crater formation on a magnesium oxide coated plate) faded out around  $10\mu$ . If, however, the bubble burst occurs on salt water, the evaporating droplets leave salt nuclei behind them, and these can behave as condensation nuclei in a cloud chamber. Making use of this, Mason (1954) detected between 100 and 200 nuclei per burst from a 3 mm diameter bubble;

the amount of salt per nucleus indicated a droplet diameter of between  $0.4\ \mu$  and  $1.0\ \mu$  at formation. Mason found that a smaller bubble bursts with greater violence, so that about the same number of nuclei were formed from a bubble of only 0.5 mm in diameter. In this case the order of  $10^4$  nuclei per  $\text{cm}^2$  of bubble surface will have been produced. In the case of pinene already discussed (Harper 1953 b), the number of neutral centres per  $\text{cm}^2$  of bubble surface was also of the order of  $10^4$ , and the number of large ions formed in the same apparatus with water, though not measured, was known to be about an order of magnitude bigger. It is not stretching the evidence very much, then, to say that the disintegration of a bubble blister produces enough particles to account for the large ion concentrations found in entrainment charging, and not stretching them at all if Mason's expansion chamber failed to detect the smallest nuclei. Let us, therefore, examine the supposition that we have now identified the process responsible for the charging.

Large ions are formed by splashing, spraying and the shattering of drops, as well as by bubbling. Hochschwender (1919) obtained photographs of the shattering of drops by an air blast in which there were sudden changes of velocity, and these showed the drops being blown out into the shape of a hat, the crown being a thin film, and the rim the rest of the drop, annular in form. The disintegration of the film gave rise to very small droplets which were negatively charged when using distilled water, the annulus breaking up into positively charged droplets of larger size (Pl. II(d)). The shattering of drops, then, does involve the process deemed responsible for the formation of large ions. Newitt *et al.* (1954) found that drops blown off baffles were torn away as thin sheets if the velocity were sufficiently high, whereas at lower velocities a filament was formed, and this broke up into droplets as shown in Plate II, which also shows the way in which a thin liquid lamina, produced by spraying, breaks up into droplets (Dombrowski and Fraser 1954). The process may begin with the formation of a number of perforations which expand until a lace-like structure is formed. The threads of the lace are unstable, and break up into droplets of different size, including small satellites (Plateau's spherules), also shown in Plate II. Under different conditions threads may be torn off the edge of the lamina, or, subjected to sufficiently violent disturbance, the film may disintegrate directly into droplets (Plate II). Worthington (1895) photographed the splash of a drop, finding that a cylindrical sheet of liquid was ejected, the rim of which disintegrated into drops (Plate II). Aganin (1914) followed the charging due to a jet of water impinging on a water surface as the height of the jet was varied, and concluded that the variation paralleled the number of very small air bubbles produced. The air bubbles, of course, give rise to jets and blisters, and hence also to droplet formation. We have now considered splashing, spraying, bubbling and the shattering of drops, and have found that the formation and disintegration of thin films is a process that occurs in all four cases, i.e. whenever entrainment charging is observed.



We have also seen that the disintegration of the film probably produces as many particles as large ions observed. There are, moreover, enough electrolytic ions in the liquid to charge the particles. In fact, at first sight, there seem to be too many. We estimated that a thickness of liquid of a mere  $2\text{ \AA}$  would provide enough ions in the case of pinene. Hochschwender (Lenard 1921) obtained  $5 \times 10^6$  elementary units of charge from a drop 5 mm in diameter. Nolan (1914) had previously obtained almost the same value. Again, a thickness of liquid equal to a molecular diameter would easily have supplied sufficient ions, since pure water contains  $5 \times 10^{13}$  pairs of ions per  $\text{cm}^3$  at room temperature, and distilled water the order of 100 times as many. If it were possible for a liquid film to thin down to monomolecular dimensions before disintegrating, a surface layer could be separated and provide ions to be transferred to the gas, but, in fact, any real film will disintegrate much earlier, in which case how can we get both the order of magnitude of net charge observed, and all the ions of one sign as reported for some liquids by Coehn and Mozer (1914)? Let us enquire first at what thickness the film really does disintegrate.

The surface of water must be greatly distorted by thermal agitation: Raman and Ramdas (1927) found from polarization measurements on a reflected light beam that the transition layer so formed had a thickness of  $5\text{ \AA}$ . Remembering that a film has two surfaces, we may be sure that it will disintegrate before it has thinned to  $10\text{ \AA}$ . It is well known that the survival of a water film is encouraged by the presence of a monomolecular layer of soap, which, in the case of oleate, is  $11\text{ \AA}$  thick. Johonnot (1899), using the Michelson interferometer, found that the thinnest 'second black' film of 'water' he could obtain in this way had a thickness of  $60\text{ \AA}$ . Without the use of soap to stabilize the film, it is hardly likely that it could thin to less than  $100\text{ \AA}$ , and might well disintegrate at  $1000\text{ \AA}$ . Newitt *et al.* (1954) found the lifetime of a bubble to be of the order of  $1/100$  sec, whereas the bubble burst was over in microseconds. This excludes the possibility of the blister thinning so quickly that it has insufficient time to disintegrate though ready to do so. They also collected enough debris from the bursting of the bubble to account for a film thickness of  $100\text{ \AA}$ ; the actual thickness must have been greater, because of evaporation. Now, if a film disintegrates before it has thinned to  $100\text{ \AA}$ , and perhaps a lot sooner, the resulting particles can only contain a marked excess of ions of one sign if the surface double layer extends at least to a depth of  $100\text{ \AA}$ .

Evidence regarding the double layer at a pure water/air interface is conflicting. According to Davies and Rideal (1955), figures for the resulting potential discontinuity range from  $-0.5\text{ v}$  to  $+0.4\text{ v}$ . It seems to the present author, however, that an improved picture can be obtained by considering the consequences of hydrogen bonding between water molecules (Latimer and Rodebush 1920), the existence of which is now generally accepted (Owston 1951). A water molecule in the vapour

consists of two protons each at a distance of  $0.958 \text{ \AA}$  from the oxygen nucleus, and subtending an angle of  $104^\circ 31'$  there. The bonding is covalent, the dissociation energy of the bond being  $4.8 \text{ ev}$ . The resulting dipole moment is  $1.84 \text{ e.s.u.}$  In water (and in ice) the angle between the bonds tends to become the tetrahedral angle ( $109^\circ 30'$ ), and the bond length slightly longer, but the covalent bonds are not very much changed. The two other tetrahedral directions, however, assume a new importance in the presence of other water molecules, being able to attach hydrogen belonging to another molecule. The hydrogen bond so formed is weak, the dissociation energy being  $0.24 \text{ ev}$ , but, since this is about  $10kT$  at room temperature, liquid water shows a strong tendency to form molecular aggregates. It has, indeed, been suggested, that liquid water at moderate temperatures contains quite a lot of 'ice'. The quantum-mechanical origin of the hydrogen bond is uncertain (Kellner 1952), but we can proceed without knowing what it is. It is sufficient that the tendency to attach to another molecule is directional with respect to the dipole moment of the molecule, and that energy of amount  $0.24 \text{ ev}$  is released when attachment occurs.

Molecules in the bulk of the liquid form polymeric groups, but there is no preferred orientation for them. Molecules on the surface, however, can only attach themselves to other molecules by turning so as to present their unsatisfied hydrogen bonds towards molecules within the liquid—there are no molecules in the opposite direction. In doing so, they turn their covalent bonds outwards, because the four directions are tetrahedrally related. The covalent bonds join the oxygen to the two protons, whence it follows that the dipoles are turned positive outwards. Their tendency to this orientation will be but little disturbed by thermal agitation. When, however, some molecules have already turned, their dipoles will cooperate to produce an electrostatic field tending to prevent more molecules from turning, and the building of the orientated dipole layer will be brought to a stop when the field it produces becomes large enough. Overlooking thermal agitation, this will be when the work done in turning the dipole in the field of its neighbours is equal to the energy released by the formation of the hydrogen bond. For a preliminary estimate of the strength of the dipole layer we may average out complications such as the orientation not being strictly confined to the outermost monomolecular layer, and simplify the situation by representing the protons by a uniform positive charge distribution over a plane in the surface, with the compensating negative charge uniformly distributed over a plane just below it. It then follows immediately that the change of potential when passing through the layer, expressed in volts, is numerically equal to the work done in orientating a new dipole, expressed in electron volts, and this is  $0.24$ . Thus we predict a surface dipole layer, positive outwards, giving rise to a potential discontinuity of about a quarter of a volt. Chalmers and Pasquill (1937) claim to have measured the potential discontinuity at an air/water interface as  $0.26 \text{ v}$  *negative* outwards, but

their experimental method is open to criticism, and they did not use specially purified water.

A double layer a few molecules thick is too thin to be split by the formation of a film of thickness exceeding 100 Å, and in any case water dipoles could not be divided by this sort of mechanical process. So the question now arises as to what happens to the hydrogen and hydroxyl ions which are present in significant numbers, even in pure water. Gurney (1936) has given a detailed interpretation of the solvation energy of ions. An ion brought up to the surface of a dielectric liquid from the outside is attracted to it by the image force, and has acquired somewhat less than half the solvation energy when it is halfway into the liquid. It does not acquire the full solvation energy until it is well below the surface, and is therefore still subject to an inward force until it is well buried. In the body of the liquid, the only electrostatic forces acting on the ions are those due to other ions, but as soon as an ion approaches the surface, it is repelled from it. Thermal agitation can take it a little way in this direction, but not far, because the solvation energy of an ion is two orders of magnitude larger than  $\frac{1}{2}kT$ . The ionic concentration will fall off rapidly above the depth at which the deficiency of solvation energy is equal to  $\frac{1}{2}kT$ . The solvation energy when the ion is surrounded by a spherical shell of liquid is given by Gurney's eqn. (5), but there is a mistake in it arising from the integral not having been given the proper limits. Correcting the mistake, the decrease in solvation energy resulting from replacing liquid outside a sphere of radius  $c$  by vacuum, when the charge on the ion is  $q$  and the dielectric constant  $K$ , is  $q^2(K-1)/2cK \sim q^2/2c$  for large values of  $K$ . The decrease will be less, by more than a factor 2, for an ion at a distance  $c$  from the surface, because more than half the space formerly replaced by vacuum is now occupied by liquid. For an order of magnitude consideration we shall take the factor as 3, and so find that the deficiency of solvation energy is equal to  $\frac{1}{2}kT$  for  $c \sim 200$  Å. Few ions will approach the surface much closer than this, and the surface dipole layer is too far away to affect the ions at this depth. Observing that the deficiency of solvation energy is the same for all singly charged ions, since it comes from lack of dielectric polarization at a relatively great distance from the ion, we find no reason to suppose the negative ions to behave differently from the positive, and hence no reason for an ionic double layer to form.

The conclusions we have reached about the surface of pure water are as follows. An outer layer, about two molecules thick, is ill-defined because of thermal agitation. The water molecules are partially orientated positive outwards within a layer a few molecules thick, the difference in potential arising from this double layer being about a quarter of a volt. A layer 100 Å and more thick is denuded of ions, below which the concentration increases similarly for hydrogen and hydroxyl ions. If now a bubble film thins to 100 Å before disintegrating, there is no reason why pure water should give entrainment charging at all; if it disintegrates



at ten times this thickness, as we found probable, there should be ions of both signs in approximately equal numbers, not all negative as found by Coehn and Mozer (1914).

The addition of small quantities of electrolyte such as NaCl to pure water will not make any fundamental change in our picture of the surface, but it is quite otherwise if surface active electrolytes such as, say, sodium oleate are present. The distinction is between capillary active and capillary inactive solutes. Organization of the long-chain molecules into a surface film represents a state of lower energy than solution in the water ; when the surface film dissociates into ions, the long-chain ions can remain bound to the surface. The ions of opposite sign will then form a space charge within the liquid, thus giving rise to a double layer the thickness of which is given by Debye-Hückel theory (Butler 1951) as  $3C^{-1/2}$  Å, where  $C$  is the concentration in mols/litre. The thickness is 100 Å at  $C \sim 10^{-3}$ , and 1000 Å at  $C \sim 10^{-5}$ . We have, however, ignored the repulsion of the ions (i.e. of the space charge) from the surface ; the repulsion will be important at  $C \sim 10^{-3}$ , but not at  $C \sim 10^{-5}$ . If entrainment charging is due to the separation of a surface film of thickness between 100 and 1000 Å, then we may anticipate a critical dependence of charging on concentration at very low concentrations, for the case we have just been considering. Such a critical dependence on surface active contamination, and traces of electrolyte, is well known from experiments too numerous to quote. To examine them in detail in the light of these ideas would be too speculative.

It has been assumed so far that the division of a double layer by the formation and disintegration of a thin film will result in the smaller particles carrying a charge derived from the *outer* part of the double layer, the film having become too thin to accommodate the inner part, because of the proximity of the two sides of the film. This need not necessarily be so. If we start with a surface active layer say positively charged, and a negative space charge below it, the stretching of the film as it thins may cause the surface skin to tear, so that the thinnest parts of the film are formed where there are holes in the skin, in which case these parts will contain negative ions only. After disintegration, the smallest particles would then carry a charge derived from the *inner* part of the double layer.

The orthodox explanation of entrainment charging has been examined *in extenso*, because it now seems possible to state that it can only be maintained if made dependent on the presence of surface active electrolytes. This would force the conclusion that most experiments on entrainment charging have been vitiated by the presence of contamination, but would account for the discrepancies between different observers, and these are characteristic of the subject. For example, Eve (1907) investigated a variety of liquids by spraying, and obtained the largest charging with ether, whereas Coehn and Mozer (1914) obtained the smallest charging with ether.

An alternative explanation of entrainment charging can, however, be suggested, not dependent on contamination. The orthodox explanation involves selective sampling of an equilibrium distribution of ions, the rate at which this happens being immaterial (except for the hypothesis of reversed charging). We now recall that, in modifying the theory of statistical fluctuation charging, it was necessary to consider whether ionic conductivity might be sufficient to permit a redistribution of charge during the formation of the droplet, but that we rejected the hypothesis of droplet charging by electrostatic induction, except as a disturbing factor. There is, however, another way in which random sampling may be interfered with by the dynamics of film disintegration. The ultimate stage in the disintegration is generally the breaking up of a thread of liquid into droplets, and it is usually supposed that each droplet is accompanied by a very small satellite or Plateau's spherule. Let us suppose that it is the satellite which will evaporate to form a neutral centre if uncharged, or a large ion if it contains one or more electrolytic ions at birth. If this is a correct picture, then the large ions arise from those portions of the liquid which are 'satellites-to-be', *and which happen, just before break-up, to contain electrolytic ions.* Thinking of pure water, which contains one pair of ions, on the average, in a sphere of diameter  $0.3\mu$ , we would not expect 'satellites-to-be' to contain many ions; we would expect, rather, that the 'large-ions-to-be' are those 'satellites-to-be' *which happen to contain more than an average allowance of ions.* Assuming this to be so, consider what happens when the satellite is in process of becoming detached from its parent drop. Though it contains more than its fair share of ions, there is no reason at the outset why these should be of one sign rather than the other. If they represent an improbably high concentration of ions, they will have a strong tendency to disperse because of diffusion and mutual repulsion, but have a limited time in which to do so—the time it takes the satellite to form. Now the ions of greater mobility are more likely to disperse into the parent drop, and escape retention by the satellite. This will lead to the formation of an excess of large ions containing that sign of electrolytic ion which has the lesser mobility, negative in the case of water. It must be admitted that this mechanism could hardly account for *all* the ions being of one sign, though it might well bias the sampling in favour of that sign.

The problem of entrainment charging is twofold, involving both the mechanism of droplet formation, and the question of the surface double layer. We have seen that it is only one kind of double layer, or more generally one component of a complex double layer, that is able to give rise to entrainment charging: a double layer in which one-half is bound to the surface, and the other half a space-charge within the liquid. This is precisely the component that gives rise to electrophoresis. In the case of air bubbles in water the surface active ions are, so to speak, attached to the air bubble, the ions of opposite sign being free within the water. The bubble therefore moves when an electric field is applied,

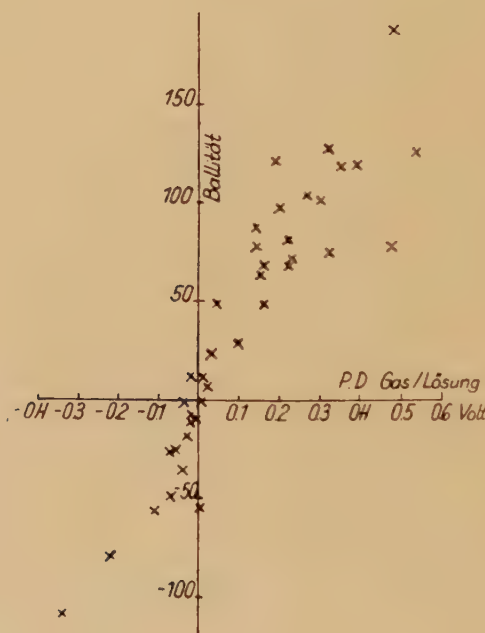
and from its velocity one may deduce the strength of the relevant component of the double layer, or zeta-potential as it is called. Alty (1924) concluded that the zeta-potential would probably be zero in perfectly pure water, but Currie and Alty (1929), working with more highly purified water, found they could still observe, and measure, electrophoretic effects. These effects, however, were very small, and cannot be interpreted in terms of a simple zeta-potential, which, in any case, could not have amounted to even as much as a few millivolts. The electrophoresis depended on the rate at which the bubble was dissolving in the water, and may have had nothing to do with a true equilibrium condition. In the case of most solutions investigated by Currie and Alty, it took over half an hour for the bubble to stabilize; this suggests that the effect was at least partly due to contamination. The experiments should almost certainly be interpreted as confirming that there is no significant double layer of the type that gives rise to a zeta-potential at the interface between air and pure water. This conclusion is compatible with the existence of a dipole double layer, which does not give rise to a zeta-potential.

What has been said about pure water makes it very difficult to interpret an important paper of Frumkin and Obrutzhewa (1931) which is mainly concerned with correlating some early measurements of Christiansen's on spraying with their own measurements on surface potential. Concentrated aqueous solutions were used containing two solutes of different character. One was a capillary active non-electrolyte giving rise to a surface layer of orientated dipoles, the potential difference through it being measured virtually in the absence of electrolyte and relative to pure water for which the surface potential was arbitrarily taken as zero. Some surface layers were positive outwards, and some negative. It was found that the addition of a second solute, a capillary-inactive electrolyte, had little effect on the measured surface potentials, as would follow from the existence of a surface layer denuded of ions. When, however, such a solution was sprayed, it appeared that with the dipole layer positive outwards negative ions were transferred to the gas, and conversely. This at once proves that the spray charging cannot be the result of stripping off the outer half of the dipole layer, because the wrong sign of charge would be predicted; the ions found in the gas must have been electrolytic ions in the solution. It suggests that sufficient ions to give the charging had been held in the dipole layer electrostatically, though not enough to alter the surface potential appreciably, and far fewer than in the bulk of the liquid, because of the Boltzmann factor. If this were so, one would expect the charging to vary exponentially with the surface potential, because of the Boltzmann factor. Figure 5, however, shows that experiment gave a nearly linear trend when charging was plotted against surface potential for different substances. When plotted against surface potential for different concentrations of a weakly capillary-active solute, ethyl acetate gave proportionality, but propionic



acid did not. It should, however, be remarked, that these capillary-active substances can hardly be classified as non-electrolytes. Frumkin and Obrutzchewa found that dilute solutions of electrolyte showing inversion with varying concentration had the inversion suppressed when a capillary-active polar solute was added; in accordance with former observations, the sign of charging at all concentrations was determined by the sign of the surface potential. But results with capillary-active electrolytes did not always fall into line. Nevertheless, it would be much the most natural explanation of the facts to assume that the primary effect is one in which the dipole layer is loaded with electrolytic ions of sign opposite to that of the outer ends of the dipoles, were it not for the difficulty of the Boltzmann factor. We shall therefore examine this factor rather more carefully.

Fig. 5



From *Kolloid Zeitung*.

Dependence of charging (Ballität) on surface potential.

We have seen (Gurney 1936) that an electrolytic ion loses solvation energy as it approaches the surface, a simplified picture of the process making it lose one-half by the time it reaches the surface if the dielectric constant is large. This assumes that the surface remains flat, the ion merely replacing a surface molecule. The solvent molecules surrounding the ion are, however, loosely bound to it, and some will be pushed above

the surface as the ion comes near it, forming a pimple. The pimple enables the ion to retain some solvation energy that would otherwise be lost; the larger the pimple the less will be lost. The tendency for the pimple to grow bigger is opposed by the work required to create fresh surface, and this will clearly set a limit to the size of the pimple. Let its radius during growth be  $r$ , its equilibrium radius be  $r_0$ , and let the radius of the ion be  $a$ , with charge  $q$ . If the radius of the pimple increases from  $a$  to  $r$ , solvation energy of amount  $q^2(1-a/r)/4a$  is gained, but an amount of work equal to  $2\pi T(r^2-a^2)$  is done in the process, where  $T$  is the surface tension. The equilibrium radius is given by

$$\frac{d}{dr} \left\{ \frac{q^2}{4a} \left( 1 - \frac{a}{r} \right) - 2\pi T(r^2 - a^2) \right\} = 0,$$

whence  $r_0^3 = q^2/16\pi T$ . Putting in numerical values for a singly charged ion in water gives  $r_0 = 4.0 \text{ \AA}$ , and if we take  $a = 1.6 \text{ \AA}$ , we find that the energy gained is  $0.3 \times \text{solvation energy} - 0.38 \text{ ev}$ . Thus, instead of requiring half the solvation energy to bring an ion to the surface, pimple formation reduces it to  $0.2 \times \text{solvation energy} + 0.38 \text{ ev}$ , which, for  $a = 1.6 \text{ \AA}$  amounts to  $1.28 \text{ ev}$  instead of  $2.25 \text{ ev}$ , a saving of approximately  $1 \text{ ev}$ . Nevertheless, even for negative ions in water, which can acquire up to  $0.24 \text{ ev}$  from the double layer, there remains approximately  $1 \text{ ev}$  to be overcome by thermal agitation before the ion can get to the surface. The Boltzmann factor is  $e^{-40}$  or  $10^{-17}$ , so it still seems impossible for there to be a significant concentration of ions on the surface, unless they are bound to surface active material.

### 2.2.2. Formation of large ions by entrainment: conclusions

Bubbling, splashing, spraying and the shattering of drops are the processes which give rise to entrainment charging. All of these processes involve the formation of sheets of liquid which thin down to between  $100 \text{ \AA}$  and  $1000 \text{ \AA}$  before disintegrating. Depending on circumstances, these either break up directly into pieces which contract into drops, or, more usually, develop perforations which enlarge until a lace-like structure is formed. The threads of this structure immediately disintegrate into small drops. These drops are accompanied by much smaller satellites. We have seen that this mechanical disintegration of the liquid results in entrainment charging by the direct transfer of electrolytic ions into the gas, random sampling of statistical fluctuations in the distribution of the ions explaining the charging in the case of droplets of diameter around a micron from sufficiently non-conducting liquids. There seems no reason to doubt that, when large ions are observed, these come from the evaporation of the smallest particles formed by the same mechanical disintegration of the liquid, but the sampling of the electrolytic ions in the liquid is only random when conditions are such that large ions will not be formed in significant numbers. Differing explanations of the way in which large ions acquired their charge have been reviewed,

according with differing views concerning the departure of the sampling from random. Six causes of departure from random sampling have been considered, and can now be assessed as follows.

(i) Charging by electrostatic induction in the field of an extraneous contact potential. This cannot normally be important.

(ii) Charging by electrostatic induction in the field of drops already charged. This is a secondary effect.

(iii) Partial (or even complete) separation of a double layer into its two components. This has become the orthodox explanation, but we have seen that it calls for an *ionic* double layer in which one sign of ion is bound to the surface, the other forming a space charge within the liquid. Pure water cannot possess such a double layer, though hydrogen bonding between molecules will cause the dipoles to be partially orientated in a surface layer a few molecules thick, positive outwards, and give rise to a surface potential of about a quarter of a volt. To a depth of 100 Å or more there is a layer which is denuded of electrolytic ions. If now traces of surface active electrolyte are added to the water, giving rise to a charged skin, the double layer thus formed *can* give rise to entrainment charging, but it is uncertain whether the large ions will carry excess charge from the surface skin, or from the space-charge below it. The orthodox explanation of entrainment charging seems only tenable if it is assumed that most experiments have been unintentionally affected by the presence of surface active contamination.

(iv) If entrainment charging results from the separation of the two components of a double layer, with the preferential transfer of ions of one sign into the gas, then immediate repetition of the process on liquid already denuded of ions of this sign will give rise to preferential transfer of ions of the opposite sign. This is a secondary effect.

(v) Doubts thrown on the orthodox explanation of entrainment charging suggest that a new hypothesis may provide the correct explanation. In this hypothesis the significant process is the formation of a satellite droplet from its parent drop. Redistribution of statistical fluctuations of charge during the formation of the satellite result in the satellite tending to retain ions of low mobility, and the parent drop ions of high mobility.

(vi) Particulate matter present in the liquid is likely to adsorb ions preferentially, and convey them into the gas. This possibility is bound up with the existence of 'veils' now to be described.

### 2.3. *Motes and Veils*

It is difficult to obtain a liquid free from suspended particles, particularly in the case of water; conductivity water normally contains motes. Dorsey (1940, p. 318) gives a list of references. One method of eliminating motes is distillation *in vacuo* with avoidance of bumping and bubbling, but, when the receiver containing the dust-free water is shaken,



it is almost certain that the water will again become contaminated with motes, and it is no easy matter to avoid this trouble. No experiments on entrainment charging have even attempted to do so, and, since motes inevitably provide neutral centres or large ions, the neglect may be important. Curry and Alty (1929) noticed in their electrophoresis experiments, in which they followed the charge on a slowly dissolving bubble of air in water, that the charge generally decreased discontinuously, by the order of  $10^{-4}$  e.s.u. Curry and Alty suggested that instabilities occur as the surface concentration of ions increases, but it is surely more likely that the sudden decrease of charge represents the loss of particulate matter.

Insoluble impurities need not necessarily be particulate in the sense of being motes. Contamination *may* be the disperse phase, but it is also possible for the liquid to be the disperse phase. In the case of water, this has been recognized for some time (Dorsey 1940, p. 401). Ice has a grain structure: melting always begins at the grain boundaries. This points to the existence of an intercrystalline material, which might consist of a concentrate of traces of impurity originally dissolved in, or dispersed in, the water. The material seems always to be present, and Quincke (Dorsey 1940, p. 401) suggested that the network of intercrystalline material remains in the water after melting, giving water a foam-like structure. For forty years the suggestion was looked upon as fanciful—as is hardly remarkable. What is remarkable in the extreme is that the supposed intercrystalline network has recently (Hast 1952) been transferred from the water to a vacuum with quite good retention of its original shape, and photographed with an electron-microscope. Under ordinary conditions the network doubtless becomes broken up soon after melting has occurred; the shape of this thin veil-like structure can only be preserved by employing a special technique. It nevertheless seems likely that a water *surface* will be covered by a veil, and that it will have an effect on entrainment charging. The discontinuous decreases of charge on a dissolving air-bubble, observed by Alty, would then have been due to the folding of a veil, or the detachment of a part of it, and the veil must have been charged, as one might expect.

Veils derived from crystal grains are of quite general occurrence. Hast obtained them by dissolving NaCl in water, by dissolving marble in acids, and even by the condensation of water vapour followed by evaporation. Veils contain extremely little material, and are difficult to photograph, but, as may be seen from Plate III, their existence seems demonstrated. It appears that, while forming, they follow the contour of the growing crystal, but, during the dissolving of the crystal, they are left behind unchanged. They appear to have a fibrous, mat-like structure, which suggests that they are built out of contamination, though Hast does not exclude the possibility that they have to do with the natural mechanism of growth of the crystal. More will be said about this possibility later § 5.3.2.

## § 3. LIQUID/SOLID CHARGING

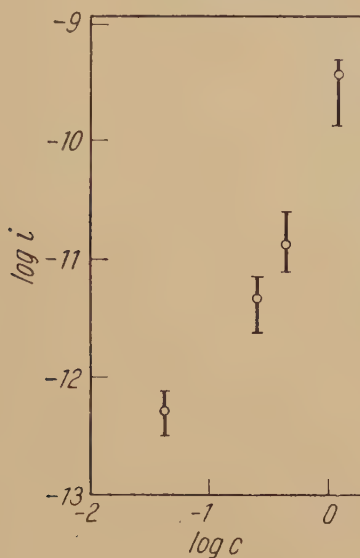
3.1. *Charging of a Liquid by Streaming*

If we accept the orthodox explanation of gas/liquid entrainment charging, in which a surface double layer of charge is disrupted so that one sign of charge is transferred preferentially into the gas, then we must postulate a double layer in which ions of one sign are bound to the surface, ions of the other sign forming a space charge within the liquid. If, now, we have a liquid/solid interface, ions may be bound to the solid surface by adsorption, and if the liquid moves tangentially to the surface, ions of opposite sign may be removed to a distance, so giving rise to a separation of charge. A potential will continue to be built up until the conduction current back through the liquid just balances the convection current due to the flow. Such streaming potentials, discovered by Quincke, and explained by Helmholtz, are well known in electrochemistry (Butler 1951). For ordinary electrolytes they are quite small, but for nearly insulating liquids such as petrol they may be dangerously large, and unless precautions are taken against them when emptying oil tankers, explosions are liable to occur. Voltages generated in this way are a maximum for a bulk electrical resistivity which is high enough to permit the voltage to build up, but not so high that there are insufficient ions to generate much charge. Industrial liquids usually thought of as non-conducting, such as fuel oils and organic solvents, provide examples. It must, however, be remembered, that liquids such as alcohol, sufficiently conducting to prevent the generation of dangerous charges by streaming, may nevertheless do so by entrainment charging; there are plenty of ions to generate the charge, and it cannot leak back through the gas phase by conduction, but may do so by sparking.

It has only recently been demonstrated that the generation of static by flow through pipes is a streaming phenomenon, and not, say, due to the presence of particulate matter, as is the case in the charging of a gas flowing through an orifice (Wesendonck 1892). Nederbragt drew attention to the importance of impurities in hydrocarbons flowing through pipes, in the discussion on the paper in which Harper (1953 b) reported that their presence was essential for entrainment charging to occur with these liquids. Nederbragt found that small quantities of triethanolamine oleate and magnesium petroleum sulphonate, which are slightly ionized in hydrocarbon solutions, and are also surface active materials, increased the charging by an order of magnitude, when benzene was forced through a steel capillary. Strong charging was found at a resistivity of  $10^{11}$  ohm cm. Gill (1954) investigated the streaming charging of mixtures of alcohol with industrial petrols, and found that the charging disappeared below a resistivity of about  $5 \times 10^9$  ohm cm. Hampel and Luther (1955) have carried out experiments in which polar substances were added to purified *n*-heptane. The heptane had a bulk resistivity exceeding  $10^{17}$  ohm cm, and gave no charging in an apparatus capable

of detecting  $10^{-14}$  amp. A freshly deoxidized platinum capillary was used, water and oxygen being excluded from the liquid by performing the experiment in an atmosphere of nitrogen. In general, liquids that had not been specially purified gave the less charging the lower their conductivity. Progressive addition of polar substances to heptane gave

Fig. 6



From *Naturwissenschaften*.

Effect on charging of *n*-heptane of adding ethyl alcohol.  $c$  in mol/litre,  $i$  in amps.

#### Electrostatic charging of *n*-heptane with additives

Additive	Concentration in mol/litre	Conductivity in $\text{ohm}^{-1} \text{cm}^{-1}$	Current in amps
Butyl acetate	0.1	$3 \times 10^{-16}$	$+4 \times 10^{-13}$
Ethyl alcohol	0.1	$6 \times 10^{-16}$	$+2 \times 10^{-12}$
Ethylamine	0.1	$3 \times 10^{-15}$	$-6 \times 10^{-12}$
Acetic acid	0.1	$1 \times 10^{-16}$	$+2 \times 10^{-13}$
Butyl stearate	0.006	$1 \times 10^{-17}$	$+2 \times 10^{-12}$
Octadecanol	0.007	$7 \times 10^{-17}$	$+8 \times 10^{-13}$
Octadecylamine	0.004	$1 \times 10^{-14}$	$-5 \times 10^{-10}$
Stearic acid	0.021	$4 \times 10^{-16}$	$+4 \times 10^{-12}$

an initial increase in current, followed by a decrease; *n*-heptane containing ethanol acquired a positive charge. Quantitative results are given in fig. 6 and the table.

Recent evidence, then, indicates that we may accept a theoretical treatment in terms of streaming potential. Such a treatment was given



by Cooper (1953). He derives the following formula for the transport current in a pipe of diameter  $d$  in which the mean velocity is  $\bar{u}$ :

$$I = \frac{\zeta K \bar{u}}{4} \left( \frac{\bar{u} w d}{\eta} \right) \left( \frac{\rho}{w \bar{u}^2} \right).$$

The liquid has density  $w$ , coefficient of viscosity  $\eta$ , dielectric constant  $K$ , and sets up an electrokinetic potential  $\zeta$  at the wall of the pipe.  $\rho$  is the viscous drag per unit area. The units are electrostatic, and c.g.s. mechanical. The equation is written in the above form to include the case in which the flow is not streamline, when  $\rho w \bar{u}^2$  is a function of  $\bar{u} w d / \eta$ , to be determined by experiment as discussed in treatises on hydrodynamics. For streamline flow the equation reduces to  $I = 2 \bar{u} \zeta K$ .

Cooper points out that although, in practice, a transport current usually gives trouble because of the liquid leaving the pipe, it may also generate localized voltages within the pipe if the lines of flow are distorted by projections from the walls. He furthermore attributes charges formed by filtration, and by the draining of dry-cleaning liquids from textiles, to the streaming of the liquid through a multitude of capillaries in parallel.

The relaxation time for the disappearance of a bulk charge in a conducting medium is  $K\rho/4\pi$ , where  $\rho$  is the resistivity in e.s.u., and is independent of the geometry of the situation, so that agitation makes no difference to the rate of disappearance of charges, though it may redistribute them to advantage.

If a potential difference  $V$  builds up between the ends of a tube in which the flow is streamline, then a conduction current flows in the direction opposite to the transport current, and the two become equal and opposite at equilibrium, whence we find

$$V = - \frac{\zeta K P \rho}{4\pi \eta}.$$

in which  $P$  is the pressure difference between the ends of the tube. This is the form in which the formula is used in electrochemistry (Butler 1951). The dimensions of the tube do not occur (explicitly) in it. The experiments on hydrocarbons give values for  $\zeta$  of the order of 0.1 v;  $\eta$  is about 0.003 poise and  $K$  about 5. In the handling of petrol a conservative figure for  $P$  will be  $10^5$  dynes/cm<sup>2</sup> (a 4 ft head). For these conditions we find, changing units appropriately, that  $V = 10^{-6} \rho$ , with  $V$  in volts, and  $\rho$  in ohm cm.  $\rho = 10^{11}$  ohm cm is therefore capable of generating 100 000  $V$ —clearly a dangerous potential, though it may be that the values of  $\zeta$  occurring in industrial practice are less than 0.1 v, in which case the generated voltage will be smaller. Nevertheless, there is reason to think that the resistivity of  $10^{11}$  ohm cm suggested by Cooper as an upper safe limit when streaming petrol, ought to be reduced by one or two orders of magnitude. It should also be emphasized that low resistivity is no protection against entrainment charging.

### 3.2. Charging by Freezing

Considerable potentials can be set up across an ice/water interface when the ice is forming, if the water contains a small quantity of electrolyte. The effect was discovered by Workman and Reynolds (1950), and has also been investigated by Gill (1953 b). Some of the experimental findings are disputed (Gill and Alfrey 1952 b, Workman and Reynolds 1952), but the mechanism involved appears to have been established. It is supposed that one sign of ion is preferentially frozen into the ice to form temporarily charged impurity centres, the other sign being left behind as a space-charge in the solution. This gives rise to an open-circuit potential which may amount to 230 v when the ion being frozen-in is ammonium. On closed circuit, currents of the order of a microamp are obtainable; for sufficiently low values of external resistance the current, according to Gill, is independent of the resistance, but proportional to the rate of freezing, in agreement with the presumed origin of the effect, this being a mechanical transportation of charge. There is no effect with pure water, nor if too much electrolyte is present. The optimum concentration is in the region of  $10^{-6}$  normal for ammonium hydroxide, and  $10^{-5}$  normal for  $\text{KMnO}_4$ . Gill supposes that, when the concentration is too high, so many impurity centres are frozen into the ice that it becomes sufficiently conducting to short circuit the source of potential. Such ice is, in fact, found to have relatively high conductivity, but Workman and Reynolds rather unexpectedly find that the open circuit potential is independent of the rate of freezing, which is difficult to explain. It is also difficult to explain their observation that the effect disappears with a high degree of supercooling. Nevertheless, the correctness of some explanation in terms of the preferential freezing-in of ions seems inevitable because of the outstanding ability of ammonium and fluoride ions to generate charge: these two ions have an inert gas structure similar to the water molecule, and can therefore easily enter into the ice lattice. The maximum charging observed was

530 000 e.s.u./cm <sup>3</sup> (ice negative)	from NaF
260 000 e.s.u./cm <sup>3</sup> (ice positive)	from $\text{NH}_4\text{OH}$ .

### 3.3. Thunderstorms: charging by rime formation

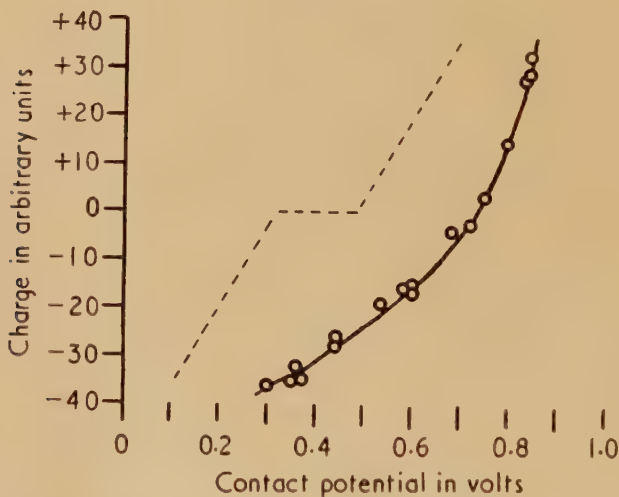
Though rather outside the scope of this article, the magnificent electrostatic displays provided by nature in the form of thunderstorms must be mentioned, if only because they involve charge generation on so enormous a scale. It is of great interest to enquire what physical process is responsible for the charge generation; it is also of great practical value to find out. The tremendous voltages that are developed present no problem; it is agreed that the separation of charge that causes them is brought about by the transport of small positively charged particles by the updraught within the thundercloud to the upper regions, the compensating

negative charge falling through the cloud on the relatively large precipitation elements. It has not yet been established, however, by what process the precipitation elements acquire their negative charge: at least nine theories have been proposed. The alternatives have been reviewed by Mason (1953 a, b) who rejects earlier theories on the grounds that either they are not in conformity with the known facts about thunderstorms, or are incapable of accounting for sufficient charge separation. He concludes that the only theory capable of explaining the facts is that the charge generation takes place when *supercooled* water droplets impinge on an already frozen water surface, themselves becoming frozen in the process to form rime. It is known from laboratory experiments that this process generates considerable negative charge, the compensating positive charge being supposed to be carried away on ice spicules.

#### § 4. MERCURY

Being both a liquid and a metal, mercury would seem to be specially provided by nature to form a link between liquid solid charging, and metal/insulator charging. Since, furthermore, friction as ordinarily understood cannot be involved in the charging, such charging as does

Fig. 7

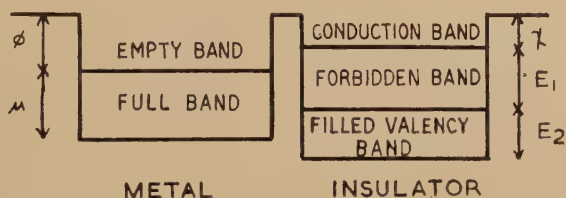


occur must be, in some sense, contact charging. Medley (1953 a) obtained charge densities up to 500 e.s.u./cm<sup>2</sup> by working *in vacuo*, and earthing the other side of the thin insulating film charged by the mercury, thereby discouraging discharging. 500 e.s.u./cm<sup>2</sup> is greatly in excess of charge densities normally encountered; here then is proof that mere contact is capable of generating *static charges of sufficient magnitude* to account for 'triboelectrification'.



Perucca (1921 a, b, c) found that the contact potential of mercury against a standard metal surface changed with time as the fresh surface aged in air, and that this was correlated with a change in its ability to charge a glass surface by contact. The ageing led to an actual reversal of the sign of charging, and, when the charge is plotted against contact potential, the relation is not so far from linear that one cannot attribute the departure to a secondary effect (see fig. 7). It is tempting to interpret this observation as supporting the hypothesis according to which charging is the result of electronic transfer between the top of the Fermi distribution in the metal, and the insulator. This hypothesis of contact electrification has recently been discussed by Vick (1953). He points out that the disposition of conduction and forbidden bands in an insulator, and of the Fermi level in a metal, is normally such that electrons cannot flow from metal to insulator because there is no lower vacant level in the insulator, and cannot flow from insulator to metal because all electrons in the insulator are deeper than the unoccupied levels in the metal (see fig. 8). In an unusual case, however, it may theoretically be possible for electrons to flow from a metal straight into the conduction band of an insulator.

Fig. 8



Mott and Gurney (1940) conclude that this would happen with caesium on AgBr in the absence of any surface contamination. Could mercury/glass also be such a case? If it were, it would be necessary to suppose that ageing could lower the top of the Fermi distribution at the mercury surface sufficiently for it to be possible for electrons to pass from the full band in the glass to the mercury. But now, during the progressive lowering of the Fermi level, there must come a stage when electrons can no longer leave the metal and cannot yet leave the insulator—just as in fig. 8. In the plot of charging against contact potential, there would then be a range of contact potential for which no charging could occur, as illustrated by the dotted line in fig. 7. The experimental curve shows no such discontinuity. It follows that the observations cannot be explained in terms of a simple metal/insulator distribution of electron levels. If the charging is by electron transfer, then there must be a *partially* filled band belonging to the glass or its surface, to account for the observed charging curve. This, however, would make glass a conductor, or at least a semiconductor. It is true that the surface conductivity of glass is much greater than the volume conductivity, but, even so, one is reluctant to regard this as a case of semiconduction. The surface conductivity is very dependent on

the extent of water adsorption ; moreover the changing contact potential of the mercury must be due to some progressive adsorption, with or without chemical change, quite probably oxidation. It seems that it is to these surface layers that we must look for an explanation of the charging, perhaps even for the existence of charging. We recall that the temperamental nature of 'triboelectrification' led us to anticipate such a conclusion.

But let us return to the mercury. It is uncertain, whether a neutral gas gives entrainment charging with pure mercury, or not. Bühl (1926), following the earlier work of Becker (1909), took considerable trouble in purifying his mercury, but used air for his experiments which were carried out by the methods of splashing, bubbling and spraying. He considered spraying to give the most reliable results, because the mercury surface is constantly being renewed in this method, and he got positive ions only, when using the purest mercury. There were very few neutral centres ; the ions were shown to contain mercury. Negative ions appeared as the mercury became contaminated by the gas. All this is strong circumstantial evidence that the simple entrainment of mercury by a gas gives positive ions. It should be pointed out, however, that it is no easy matter to be sure of the purity of one's materials : Bühl mentions, for example, that  $10^{-11}$  parts of zinc give a recognizable increase in negative ions.

De Broglie (1910) found neither neutral centres, nor ions, when mercury was bubbled, and confirmed their absence by the following experiment. A stream of dry, filtered nitrogen was passed over mercury which was being shaken vigorously, then through a chamber in which a beam from an arc light was brought to a focus to show up neutral centres by scattering, and finally through an ionization chamber. At first both neutral centres and ions were apparent, but after some time *both* disappeared. They were attributed to water, and their disappearance to progressive drying.

Theory indicates that, if there is any entrainment charging of pure mercury by a neutral gas, the ions should be negative, not positive. For a metal, whose electrical conductivity is high, statistical fluctuations at satellite formation may be excluded ; the orthodox explanation of charging by the separation of a double layer is the only one possible, and it involves the existence of a contact potential between a particle in process of being torn off, and the bulk metal. This contact potential can only arise from there being a difference in work function between an assembly of a few metal atoms, and an assembly of a large number. For sufficiently few atoms one would expect this to be the case, since for one atom the work function becomes the ionization potential. Now the ionization potential is larger than the work function, so the Fermi level in a small particle will be lower than in the bulk metal. It follows that entrainment charging should give *negative* ions, if any. Once again we are forced to suspect that the experimental findings have been determined by unwanted contamination.

A critical dependence of mercury/insulator charging on the presence of traces of oxygen in the case of oxidizable insulators such as sealing-wax,

colophony and silk, was recognized by Christiansen as early as 1894. He found the insulator negative in the presence of oxygen, but positive in oxygen-free hydrogen, and supposed the negative charging to be due to the transfer of what we now call an oxygen negative ion. The effect of oxygen was more complicated when using zinc amalgam. A trace of oxygen made the insulator charge more positively—this does not occur with pure mercury; then more oxygen gave inversion. Inversion of charging resulting from the solution of metals in mercury has been investigated by Coehn and Lotz (1921), Coehn and Curs (1924), Coehn and Duhme (1924), Bühl (1926) and Medley (1953 a). It is agreed that traces of the base metals such as sodium, zinc and cadmium can give rise to inversion, whereas noble metals such as gold do not. Bühl (1926) found the zinc inversion at  $2 \times 10^{-8}$  parts of zinc in mercury. It is clear that this is no volume effect. There is only about enough zinc to form a monomolecular layer on the surface. Nor could zinc atoms be adsorbed at the surface as such. Almost certainly, they must have been adsorbed as oxide, in which case Bühl's positive ions from supposedly pure mercury may have come from the inner component of a double layer (see § 2.2.2), the outer component of which was an oxide of mercury. Bühl reports that *visible* oxide on the surface does not affect the charging, but the present author (unpublished) has found that visible and invisible oxide on solid metals can have quite different effects, and even a monomolecular layer can affect the contact potential.

We may conclude this section by noting the opinion of Christiansen (1894) that 'frictional' electrification is really due to a chemical process occurring on contact, followed by a separation of ions. We have found quite a lot of evidence pointing that way.

## § 5. SOLID/SOLID CHARGING

### 5.1. *Ionic Transfer between Solids*

The mechanisms already considered by which charge is transferred when one material concerned is a liquid or a gas, point to the possibility that transfer of ions will play an important rôle in solid/solid charging. An interesting comment is made on this by Cooke (1955). He investigated the charging of small particles that had slid down an aluminium chute. Figure 9 shows how the charge varied with chute length for zircon (an insulator), and rutile (a semiconductor). The saturation in the case of rutile is attributed to a balance between frictional charging and a discharge current through the semiconducting particles. Cooke then found that glass microspheres behaved differently according to whether they were dry, in which case they behaved like an insulator, or had adsorbed a considerable quantity of water, in which case they behaved like the rutile. But NaCl, and other salts, unless dessicated for a long period of time, gave more charging than zircon, and a similar shape of curve, in spite of experiment showing their conductivity to be higher than that of the other materials tried. It was suggested that the anomalous behaviour of NaCl and other



salts is associated with charge transfer by ions, as distinct from charge transfer by electrons.

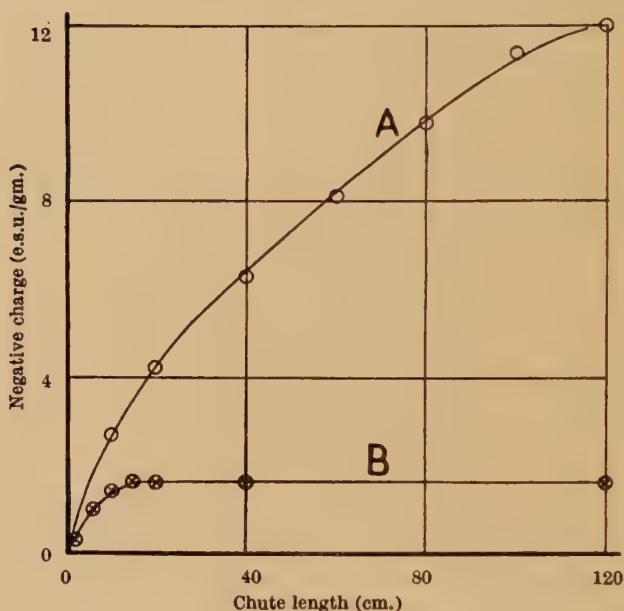
The existence of ionic transfer between solids is neatly demonstrated by some observations of Medley (1953 b), albeit for a rather special case. Polar polymers acquire electrical conductivity in a humid atmosphere due to electrolytic dissociation in adsorbed water, but only one ion is free to move, the other being part of the polymer matrix. Taking two examples, De-acidite FF' which is a strong base, and Permutit sulphonated polystyrene which is a strong acid, we have



and



Fig. 9



From Nature, Lond.

A, insulator (zircon) ; B, semiconductor (rutile). Chute aluminium.

On shaking the powders from filter paper at 30% relative humidity, Medley found that the De-acidite became consistently positive, and the Permutit consistently negative, the 'loose' ions having been left on the filter paper, and he obtained confirmatory evidence from other polar polymers.

Actually, the existence of charging by ionic transfer has been in little doubt since the work of Knoblauch (1902), though his paper has received extraordinarily little attention, and none at all from Rudge, who, in a remarkable series of experiments (1912, 1913, 1914), very difficult to interpret, nevertheless confirmed the importance of the electrochemical

properties of the materials used. Both Knoblauch and Rudge investigated the charging of acidic, basic and other powders. Knoblauch found that the sign of charging sometimes depended on whether he allowed the powder to slide off the plate by tilting it, or found it necessary to tap the plate to detach the powder, in which case he accepted the sign obtained without tapping; we note for future reference that, when tangential movement of contacting surfaces occurs the experimental findings may depend critically on the precise manner of performing the experiment. Knoblauch investigated 75 substances, and found his results compatible with one or other of two hypotheses (except in the case of diamond). The first hypothesis refers particularly to the case of a relatively soluble powder on a relatively insoluble plate (or the converse). The powder dissolves slightly in adsorbed water, and dissociates. The ions diffuse about,  $H^+$  and  $OH^-$  ions moving more swiftly than the more sluggish heavy ions. When the powder leaves the plate, some of the faster ions are left behind, these having diffused further afield. This hypothesis implies that an acid powder, providing  $H^+$  ions, should leave the plate positively charged, whereas a basic powder, which provides  $OH^-$  ions, should leave the plate negatively charged, just as was found experimentally. Knoblauch points out that a salt, which provides ions of comparable mobility, will give charging that is easily upset by disturbing factors; in fact, the charge was variable, and small. The second hypothesis refers particularly to the contact of insulators which are relatively insoluble. Ions from adsorbed water are attracted by image forces more strongly to the surface which has the greater dielectric constant. The ions of greatest mobility will be the ones that are found on that surface after separation, according to Knoblauch.

Knoblauch's two hypotheses require revision in the light of modern knowledge, but his experimental results do call for an explanation along such lines. A good deal bearing on this subject is to be found in a review by Henry (1953 b), where particular emphasis is laid on the distinction between equilibrium and kinetic effects, but his theoretical considerations are, as he admits, speculative in their application.

The hypothesis of electrolytic ions in an adsorbed water layer may appear to require too much water, but it must be remembered that a trace of hygroscopic contamination on a surface will greatly favour water adsorption, and the surfaces that one is interested in are seldom clean. There is some direct evidence on this point from an experiment of Deaglio (1928). He investigated the contact potential between silver and nickel at various pressures of moist air, and found it independent of pressure down to vacuum conditions. He then contrived, by applying a small force magnetically to a stiff spring, to vary the gap between a silver ball and a nickel ball in the range 0 to 1000 Å. For sufficiently small gaps, of the order of 100 Å, and not too low relative humidity, i.e. in excess of the order of 5%, a new phenomenon appeared, precisely as if the pair of metals with an air gap where formerly the contact potential was evident, had been replaced

by an electrolytic cell with the two metals as electrodes and electrolytic liquid filling the gap. The measured voltage of the cell was comparable with, though smaller than, the contact potential, and opposite in sign, as it should be according to this picture. There seems to be no reason why a similar layer of electrolyte should not form when a metal and an insulator, or two insulators, are brought nearly into contact, hence giving rise to electrification on separating the surfaces, perhaps by the splitting of a diffuse double layer of electrolytic ions resident in the water. Freundlich (1926) expressed himself in favour of this view.

We saw (§ 4) that there is difficulty in accounting for the charging of glass by mercury as being due to the transfer of electrons between insulator and metal, and we have just seen that cases of charging due to the transfer of electrolytic ions are quite well established. We ought therefore to consider the hypothesis that a true insulator cannot acquire a charge through the *physical* processes which occur at contact, and that, when an insulator appears to do so, it is really the presence of *electrolytic contamination* that makes the charging possible. The hypothesis can be tested by investigating whether the elimination of water or electrolyte or both from an insulator also eliminates the charging. The author (unpublished) attempted to settle this question by carefully cleaning a number of materials, paying special attention to gem stones because of their physical and chemical stability, and by testing for charging at zero relative humidity. The first combination to respond to this treatment was beryl/steel, but the experiment, though quite definite, could not be confirmed, and further inconsistencies appeared that were not resolved for a long time, for the curious reason that there are *two answers*, diametrically opposed, to the question of what happens when electrolytic contamination is eliminated. Just what this means cannot be explained until further questions have been dealt with (see § 5.3). Meanwhile it should be mentioned that Coehn and Lotz (1921) investigated the charging of pyromorphite, quartz, emery, tourmaline and carborundum against hard glass, before and after baking out in a high vacuum at red heat. Even after this treatment, which should surely have guaranteed the absence of water, charging remained. We therefore now turn to the hypothesis of electronic transfer.

### 5.2. *Theory of Electronic Transfer between Solids*

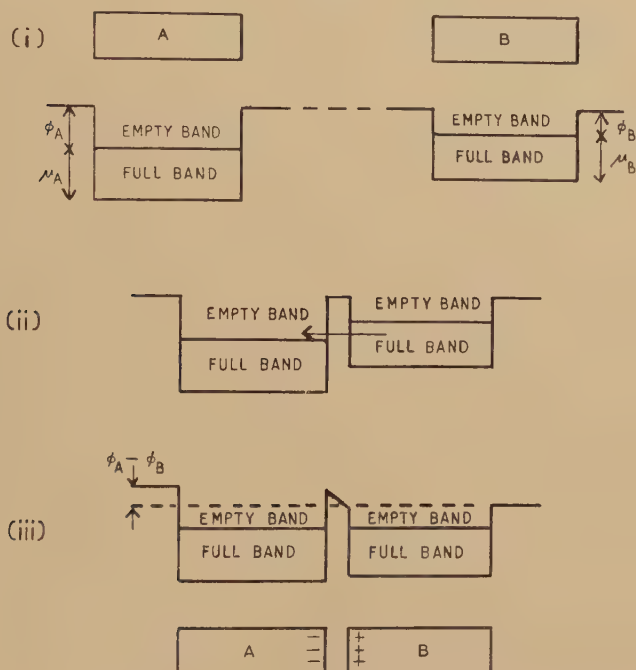
If the contact of two surfaces results in a redistribution of positive ions between them, this new equilibrium will be retained when the surfaces are separated, because, from the point of view of a positive ion, the separation is sudden. If a redistribution of electrons occurs, on the other hand, the new equilibrium will *not* be retained, because, even when the gap between the surfaces is an order of magnitude greater than the molecular diameter, electrons can leak across by 'tunneling' through the potential barrier. Furthermore, the amount of charge that leaks across depends



on the time during which the leak occurs ; the discharging is not an adiabatic equilibrium process. In principal, therefore, the final charge must depend on the rate of separation of the surfaces ; hence it is important to give careful consideration to just what *is* to be expected as a result of electronic transfer.

Harper (1951 a) has given a theory for the case of two metals, for which we can form a precise picture of the process. Referring to fig. 10 (i), A and B represent two blocks of different metal, at a considerable distance apart. The blocks being initially uncharged, the energy levels are aligned as shown. If the blocks are brought together, as in (ii), electrons from

Fig. 10



occupied levels in B will be able to pass to unoccupied levels in A by 'tunnel effect', so charging A negatively, and B positively. The charges will be mutually bound to the interface, thus forming a double layer. This double layer alters the potential distribution as shown in (iii), stopping any further flow of electrons. In this way a contact potential is established, the boundary between the two metals in contact, or near contact, being charged like a condenser by the contact potential. On separating the blocks, there is a tendency for the potential difference between them to increase as the capacitance decreases, if the blocks are insulated, but in the early stages of the separation this tendency is neutralized by a return flow of electrons through the potential barrier, and the potential difference remains constant (and equal to the contact potential) as long as such flow

is possible. Around some critical value of the gap, however, 'tunnel effect' fades out, and capacitance magnification of the potential ensues. The charge that remains is part of the charge that was originally present as a double layer astride the interface which was the seat of the contact potential. Some of this double layer leaked away, so that only part is left behind as the 'separation charge'. At no stage of the process is there any tendency for a current to flow to neutralize the double layer. There is no place for such an idea in the theory of 'triboelectrification' (Harper 1951 b). The term 'separation charging' is chosen to emphasize that true contact is not involved in the process: the charge is determined at near contact during the separation of the surfaces. To a first approximation it is the result of charging the condenser formed by the surfaces at the point of cut-off of 'tunnel effect', by the contact potential.

The theory can be made quantitative for the case of two spheres, whose radii we shall call  $R_1$  and  $R_2$  cm, the work functions being  $\phi_1$  and  $\phi_2$  volts. The current that flows between two *plane* metal surfaces, separated by a small (constant) gap, when a small voltage is applied to the gap, has been discussed in detail by Ehrenberg and Hönl (1931), using the approximation of a potential barrier of shape bounded by straight lines. The classical formula for the shape of the barrier, derived from a solution of the electrical image problem for a point charge between two conducting planes, is, however, well known to be expressible in terms of the gamma function, and Sommerfeld and Bethe (1933) have used this form to obtain, for the specific conductivity of the gap in  $\text{ohm}^{-1} \text{cm}^{-2}$ ,

$$4.0 \times 10^{10} \phi^{1/2} x^{-1} \exp \{ -0.92 x \phi^{1/2} \sqrt{(1-10/x\phi)(1-7.2/x\phi)} \},$$

in which the gap width  $x$  is expressed in Ångstrom units, and  $\phi_1 = \phi_2 = \phi$ . For the conditions in which we are interested, Ohm's law is obeyed. It is clear from the theory of potential barriers that the formula may be generalized by putting

$$\phi = \frac{4(\phi_1 + \phi_1^{1/2}\phi_2^{1/2} + \phi_2)^2}{9(\phi_1^{1/2} + \phi_2^{1/2})^2}.$$

The conductance of the gap between spheres is obtained from the formula for specific conductivity by carrying out the appropriate integration. It is found that the conductance of the gap in  $\text{ohm}^{-1}$  is

$$\sigma(x) = 2.73 \times 10^3 \frac{R_1 R_2}{x(R_1 + R_2)} \left\{ 1 - \frac{1}{0.92 x \phi^{1/2}} \right\} \\ \times \exp \left[ -0.92 x \phi^{1/2} \sqrt{\left(1 - \frac{10}{x\phi}\right) \left(1 - \frac{7.2}{x\phi}\right)} \right],$$

in which  $x$  has now become the separation of the spheres along the line of centres.

For the purposes of calculation it is assumed that a constant force is suddenly applied to one sphere to bring about the separation, the resulting acceleration being  $a$  cm/sec<sup>2</sup>. The final charge, however, proves to be so little dependent on  $a$  that we may take the formula for it as applying to the

general case with  $a$  representing the constant acceleration that would give the actual velocity in the region of cut-off. It may be seen that

$$\frac{dQ_1}{dx} \frac{dx}{dt} = -\sigma(x) V_1(x)$$

and

$$\frac{dQ_1}{dx} = V_1(x) \frac{dC_{11}}{dx} + C_{11}(x) \frac{dV_1}{dx} + (\phi_2 - \phi_1) \frac{dC_{12}}{dx},$$

whence

$$\left\{ \frac{\sigma(x)}{\sqrt{2ax}} + \frac{dC_{11}}{dx} \right\} V_1(x) + C_{11}(x) \frac{dV_1}{dx} + (\phi_2 - \phi_1) \frac{dC_{12}}{dx} = 0,$$

in which

$$-C_{12} = \frac{R_1 R_2}{R_1 + R_2} \left\{ \gamma + \frac{1}{2} \log_e \left( \frac{2R_1 R_2}{R_1 + R_2} \right) + \frac{1}{2} \log_e \left( \frac{1}{x} \right) \right\},$$

with  $\gamma$  = Euler's constant = 0.5772 (Russell 1909), and  $x$  is in cm in the formula for  $C_{12}$ .

Hence the separation charge in e.s.u. is found to be

$$\frac{(\phi_1 - \phi_2) R_1 R_2}{300(R_1 + R_2)} \left[ 1.151 \log_{10} \left( \frac{R_1 R_2}{R_1 + R_2} \right) + \zeta \left( \phi, \frac{\eta \sqrt{a}}{M} \right) \right],$$

in which

$$\zeta = 10.13 - 1.151 \log_{10} [\phi^{1/2} \{28.9 + 12.2\phi^{-1/2} - 2.5 \log_{10} (\eta \sqrt{a}/M)\}].$$

$\eta$  is an undetermined parameter which certainly lies between 1 and 100, and  $M$  is the ratio of the final potential difference to the contact potential.

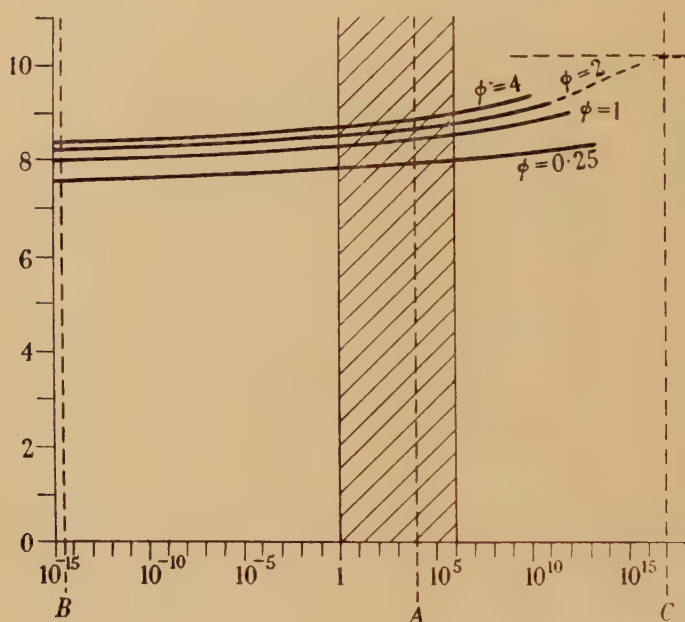
With given spheres, the only variable is the acceleration  $a$ . The formula for the separation charge then takes the form  $D \log \{E \log (Fa)\}$  in which  $D$ ,  $E$  and  $F$  are constants. Because of the doubly logarithmic form, impracticably large changes in the rate of separation would be required to show a measurable effect on the final charge. The variation of  $\zeta$  with  $\eta \sqrt{a}/M$  is shown in fig. 11. The shaded area covers the range of  $\eta \sqrt{a}/M$  that might conceivably be met with in a laboratory. The variation of the separation charge throughout the whole range is only a few units per cent. Even going to fantastic extremes does not change it a lot. We see, therefore, that, although the charging is not an equilibrium process, nevertheless the charge obtained depends so little on the rate of separation that it is not practicable to observe it. The amount of the charge, however, is determined by the order of magnitude of the accelerations that can be used in laboratory practice.

For a contact potential of half a volt, a mean work function of 5 v, and taking a reasonable value for the parameter  $\eta \sqrt{a}/M$ , the equivalent critical gap that would give the actual charge, if the cut-off were sharp, is between 12 and 13 Å. This gap is small compared with the irregularities of even a polished surface; it is therefore of some importance to enquire what is to be expected when surface irregularities are allowed for. The question will assume practical importance when comparing theory with experiment in § 5.4.



A small change in gap in the region of cut-off makes a large difference to the current, because of the exponential term in the formula for conductance, but the current is proportional to the effective area. When protuberances are present on the surface, the current will fade out at the most prominent of these, but the limitation of area carrying the current will only decrease the critical gap slightly. Regarding the actual surfaces as distortions of the ideal spherical surfaces assumed in the theory, cut-off will occur when the gap between these ideal surfaces is approximately

Fig. 11



From Proc. roy. Soc.

$\zeta$  as a function of  $\eta\sqrt{(a)}/M$ . The shaded region covers practicable laboratory conditions. A represents the conditions of the experiment. At B a separation of 1 cm is attained after  $5 \times 10^9$  years. At C the velocity of light is attained after half an Ångström unit.

equal to the (ideal) critical gap plus the height of the effective protuberance. This will alter the capacitance at cut-off. The capacitance will be further altered by the irregularities of surface, both protuberances and pits. The effect of pits on the capacitance is much less important than that of protuberances, since the former do not (in general) affect the gap between the ideal surfaces at cut-off. It is easily seen that, when the effective protuberance extends over only a small surface area, its effect on the capacitance comes almost entirely from the change in

the gap between the ideal surfaces. An approximate calculation of the capacitance at cut-off can be carried out for different distortions, by assuming the lines of force to be sufficiently parallel not to spread significantly in crossing the gap, so that the contribution to the total capacitance from the region of the gap is the sum of the capacitances of a number of plane capacitors in parallel. This will be so for the larger irregularities if they are undulating; the smaller irregularities make only a slight contribution to the total capacitance—their importance lies in altering the gap. Calculating the capacitance is a matter of algebra; the difficulty lies in assessing the importance of approximations. The author has done this for a number of special cases, from which it appears that the results can be covered by a general formula. Take quasi-Cartesian coordinates  $\xi$ ,  $\chi$ , in the surfaces, with origin at the point of contact, and let  $h$  be the departure of the actual from the idealized surface, positive when ‘proud’ of that surface. Let subscripts refer to the two surfaces, and heavy type to points on the surfaces which can become contact points. Let attention be confined to an area  $S$  in the neighbourhood of contact. Assuming that contact occurs between protuberances, i.e. that the surfaces do not fit together in such a way that pits or flats touch the opposing surface before protuberances do, the critical capacitance for the distorted surface is

$$-C_{12}'(x_c) = -C_{12}(x_c + \mathbf{h}_1 + \mathbf{h}_2 - \bar{h}_1 - \bar{h}_2),$$

where

$$\bar{h}_1 = \frac{1}{S} \iint h_1(\xi_1, \chi_1) d\xi_1 d\chi_1,$$

and similarly for  $\bar{h}_2$ . For the particular case of spheres covered with low spherical caps (nodules) of number such as could be just accommodated on a square lattice, the formula reduces to

$$-C_{12}'(x_c) = -C_{22} \left\{ x_c + \frac{\pi}{8} (\mathbf{h}_1 + \mathbf{h}_2) \right\}.$$

The error in the correction to the capacitance should not exceed 10% if  $h < 10x_c$ .

If one, or both of the materials involved is an insulator, the conception of capacitance is not directly applicable. Let us assume, in this case, that there is the possibility of electrons crossing a narrow gap by ‘tunnel effect’ in order to level up surface levels, as with Fermi levels in the case of metals, but that electrons cannot move over the surface of the insulator, since otherwise it would possess surface conductivity. Admitting that it is difficult to see why the transfer of electrons should not create a partially filled band thus permitting surface conductivity, we shall ignore this difficulty for the time being, and ask what would happen if it were overcome. The theory (Harper 1955) is much simpler than for metals. For plane surfaces, when ‘tunneling’ occurs, electrons

will stream across until the double layer so formed compensates the difference of potential between the highest occupied level on one side, and the lowest unoccupied level on the other. In the case of spheres having no surface conductivity, the process will be confined, when they are in contact, to a region of proximity bounded by the contour for which the distance between the surfaces just permits transfer by 'tunnel effect'. As the spheres are separated, this contour contracts, charge outside it can no longer cross the tunnel, but charge inside it will diminish in such a way as to maintain the 'contact potential' between the surfaces. Ultimately the area bounded by the original contour will be charged like a plane parallel condenser whose separation is equal to the critical gap, with a potential difference equal to the difference between the relevant electron levels. The capacitance so charged is easily seen to be  $R_1 R_2 / 2(R_1 + R_2)$ . *It is independent of the thickness of the critical gap.*

Skinner *et al.* (1953) also have considered the transfer of electrons from a metal to a non-metal, but Morant (1954) has drawn attention to an error in the treatment, though see also Skinner *et al.* (1954), West (1954), and Skinner (1955). The mechanism envisaged in their treatment is that of thermionic emission from the Fermi level in the metal into the conduction band of the insulator. If this can occur to a significant extent, however, the insulator would cease to insulate when tested with metal electrodes. The mechanism must therefore be a rare one, perhaps confined to intrinsic semiconductors.

A theory of the passage of electrons from a metal into a semiconductor was first given by Mott (1938), as the basis of his and Schottky's barrier-layer explanation of rectification, but, though extensive further work has been done in this field, there do not appear to be any 'triboelectric' experiments on semiconductors to which theoretical considerations can be quantitatively applied. Some qualitative implications, however, merit discussion. When two metals are brought into contact, transfer of electrons gives rise to a double layer which, to a first approximation, shifts the Fermi levels so as to be at the same depth. Since, however, the Fermi levels are slightly diffuse at temperatures above absolute zero, a small correction arises on this account. In an insulator, even though very few electrons are able to get from the 'full' band into the 'empty' band at normal temperatures, it is theoretically possible for some to do so, and the 'full' band is not really quite full, nor the 'empty' band quite empty. In consequence, when a metal is placed into contact with an insulator, equilibrium is not theoretically attained until sufficient electrons have been transferred in the requisite direction to displace the energy levels so that the Fermi level of the metal comes halfway between the top of the 'full' band of the insulator and the bottom of the 'empty' band (Fowler 1933). In practice, however, equilibrium is not attained, because the thickness of the 'double' layer required is impossibly great, owing to the maximum permitted electron density in the 'empty' band being extremely low. The insulator is not sufficiently conducting for equilibrium to be established,



not because of electrons being unable to move about, but because there are too few capable of doing so. A semiconductor is like an insulator with a less limited supply of electrons available in its conduction band, and in contact with a metal *can* attain equilibrium, with the Fermi level coming between the impurity level (for the case of an excess semiconductor) and the bottom of the conduction band. A semiconductor will therefore behave somewhat differently from an insulator possessing neither 'traps' nor surface levels, since the latter cannot charge by contact unless either the bottom of its empty band is below the Fermi level of the metal, or the top of its full band above the Fermi level, neither being likely to occur. A semiconductor, however, *is* likely to charge, since only by accident will the equilibrium configuration of levels obtain before contact, and there *are* electrons available to move in either direction. It is to be noted that the adjustment of levels due to the formation of a 'double layer' in a semiconductor requires much less charge than would be needed if both materials were metals. This follows from the fact that, though the charge on the metal is a surface charge, that in the semiconductor extends to a considerable distance within it. The field close to the metal is  $2\pi$  times the charge density, and falls off progressively in the semiconductor, not becoming zero until well within it. Thus there is a considerable distance in which the potential builds up; a smaller field, and therefore a smaller charge density, suffices, than would be required for a double layer with the opposing charges separated by molecular dimensions. When, however, the semiconductor is removed from the metal, the charge within it cannot return to the metal by 'tunnel effect' alone: it must be conducted to the surface first. For a poorly conducting material this may be impossible before the gap between the surfaces has reached the value at which 'tunnel effect' fades out, in which case the semiconductor may retain its equilibrium contact charge, and not lose part of it, as happens with a pair of metals. Nevertheless, though it is to be expected that a metal/semiconductor contact will charge, it can only be expected to do so poorly.

### 5.3. *The Two Classes of Insulators*

The actual behaviour of insulators is unexpected. Harper (1953a) found that, when examined by light contact against metals, some insulators gave abundant charging, whereas others, tested in apparatus capable of detecting one hundredth as much, gave no charging. The former may be called 'electrophilic' insulators, the latter 'electrophobic'. The division into two classes was as follows.

#### *Electrophilic insulators.*

These were hydrophilic, and were quite well cleaned for the experiment. They included two kinds of glass, magnesium oxide and vitreous silica.

*Electrophobic insulators.*

These were hydrophobic, and exceptionally good insulators, notably synthetic polymers. They were scrupulously cleaned for the experiment. Polyethylene, amber, polystyrene, nylon, silicone, polymethylmethacrylate and polytetrafluoroethylene were placed in this class in the original investigation, and cellulose, though not hydrophobic, has since been added to the list.

The question arises as to whether charge could have been lost when using electrophobic insulators because of gas discharge, but retained when using electrophilic, thus giving a false impression about the inability of electrophobic insulators to generate charge. This would be possible if an initial high charge density at the 'contact point' led to complete spontaneous discharge from electrophobic insulators before the charge density had been sufficiently reduced by surface conduction to prevent gas discharge, whereas surface conduction did prevent it from occurring in the case of electrophilic insulators. This possibility is disposed of by a later investigation (Harper 1955) in which much greater charge densities were found to be stable on very clean quartz which has a surface resistivity comparable with that of electrophobic insulators. The most likely explanation of the survival of very high charge densities (up to 8700 e.s.u./cm<sup>2</sup>) over a very small area is that the intense field to which they give rise is limited to so small a volume that there is only a negligible chance of a stray ion from radioactivity or cosmic rays appearing in this volume, so that there is nothing to start a discharge.

The further question arises as to whether electrophilic insulators would still charge if sufficiently cleaned; their hydrophilic nature suggests that the charging may be due to electrolytic contamination. The answer is that they do charge when clean (Harper 1955): the largest charges ever observed were obtained with exceptionally clean, dry, quartz surfaces. The division into two classes, then, is an intrinsic property of insulators. Furthermore, just as there are two classes of insulator, so there are two classes of contamination, the one kind, such as grease, tending to eliminate charging, the other, e.g. electrolytic contamination, tending to increase it. We now see why there are two answers to the question of what happens to the charging of an insulator when it is cleaned: an electrophilic insulator will charge more if the contamination was electrophobic, and will charge a lot when quite clean, whereas an electrophobic insulator will charge less, and not at all when uncontaminated. The cleanest electrophobic insulator so far investigated was polystyrene; an upper limit to its charging was set at four orders of magnitude lower than the enormous charging obtained using two different faces of crystalline quartz, certainly the cleanest electrophilic insulator so far used. Contaminated surfaces tend to come well within this range; scrupulous preparation of the surfaces is necessary to attain the extremes of charging.

### 5.3.1. *Electrophobic Insulators*

The existence of electrophobic insulators appears at first sight incompatible with everyday experience ; every student knows that electricity is so named because amber was the first material to become noticeably electrified, and polystyrene film is notoriously difficult to handle in the factory because it easily becomes electrified. Medley (1953 a), moreover, has obtained 500 e.s.u./cm<sup>2</sup> from Hg/polythene. How is this possible ? It occurred to the author that the difference between Medley's results and his own might be due to Medley having maintained contact for some seconds, whereas the author's apparatus made and broke contact in about 1/5 sec. The experiment was therefore repeated with contact durations up to 2 hours, but there was still no charging. The anomalous result with mercury should probably be attributed to contamination ; we have already found it necessary to explain other experiments using mercury, in this way (§ 4). The charges normally encountered on amber and polystyrene, though easily detectable, are about two orders of magnitude smaller than those found by Medley using Hg/polythene. They, also, could well be due to the presence of something other than the pure insulating material, and the reason why amber and polystyrene become more noticeably electrified than other materials be that, though not in fact easy to charge, once they have become charged, they retain their charge for a long time, even in a humid atmosphere, because of their exceptionally high surface and volume resistivities.

Neither amber/polystyrene nor amber/silica show charging : the electrophobic insulator appears to be dominant. Let us now consider what is the significance of the failure of electrophobic insulators to charge. It is clear that the full band in the insulator must be below the Fermi level of metals, so that electrons cannot leave the insulator ; the conduction band must be so far above the Fermi level that electrons cannot get into it from the metal, and there cannot be sufficient electron 'traps' in the insulator to justify its being regarded as the limiting case of a semiconductor. Furthermore, surface levels, if they exist, must be similarly situated relative to the Fermi level of metals : if empty, well above the Fermi level, if full, well below, and none partially filled around the depth at which semiconductor 'traps' might be found. The existence of electrophobic insulators is completely in accordance with theoretical views about insulators : insulators do not conduct ; how then can they charge ? This question requires an answer for electrophilic insulators.

### 5.3.2. *The Charging of Electrophilic Insulators*

To avoid the uncertainties that arise from working with multiple point contact, but yet obtain much more charging than is given by a single contact point, Peterson (1954 a) used an insulating ball rolling on a metal surface, thereby also eliminating friction. The apparatus could be baked out, and the experiments carried out under controlled conditions

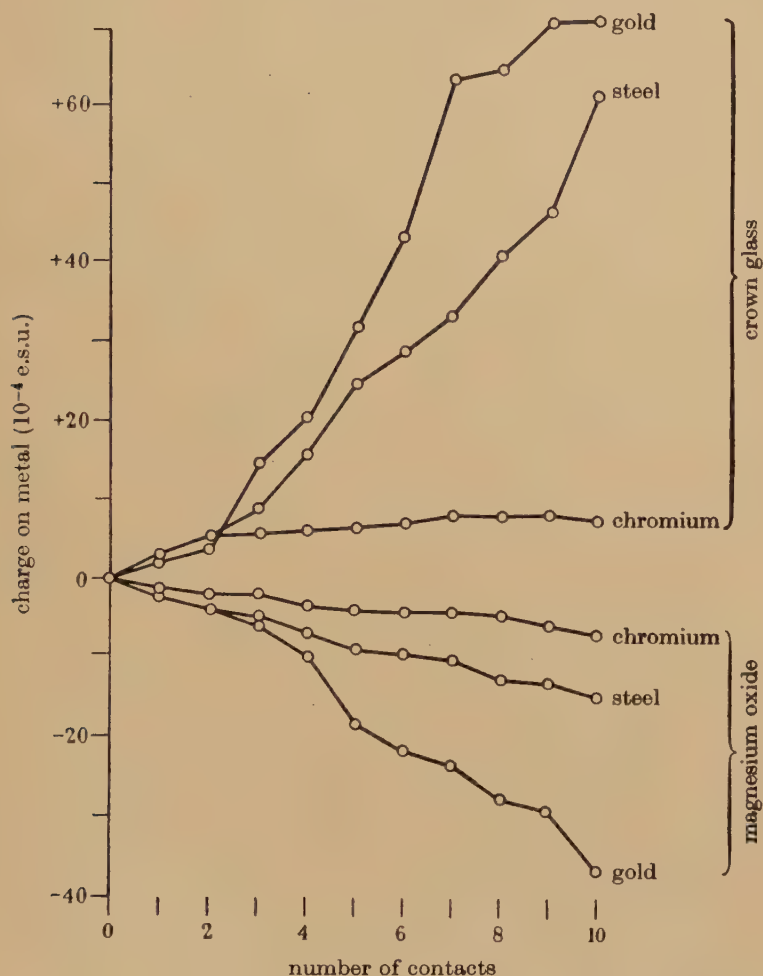


of gas pressure and humidity. The charging decreased with increasing relative humidity, disappearing at 65%. This was attributed to increased surface conductivity leading to discharging, rather than to a decrease in the charge initially transferred. Discharging through the gas was only observed after the ball had rolled a long distance, so acquiring a very large charge. It should be remarked that too little attention has been paid in the past to the question of how much charge is lost after generation, though see Medley (1950), and another paper (Medley 1953 c) on the dissipation of electrical charges generated by rollers. A related problem is the effect of an applied electric field on the appearance of static electrification. Gill and Alfrey (1949 b) supposed it to be bound up with the mechanism of charge generation, but Medley (1953 a) found that it merely gave rise to an *additional* transfer of charge. Harper (1951 b) suggested that the charge was transferred through the gas, but Peterson (1954 a) found in his experiments with a borosilicate glass ball on nickel, that the additional charge transfer took place by conduction through the adsorbed water layer. It could well be that each kind of leakage occurs under appropriate conditions. Peterson interpreted the nickel/borosilicate glass charging as being due to a transfer of electrons from the Fermi level in the nickel to a vacant level in the glass, supporting this view with the evidence that the charging altered in the proper direction (Ni less positive) after the nickel had been oxidized in a glow discharge. We have, however, found serious theoretical objections to the hypothesis of electronic transfer, and fig. 12 (from Harper 1953 a), which refers to experimental conditions rather like Peterson's, cannot be explained by any hypothesis involving a level system, because such an hypothesis would call for the three curves for magnesium oxide to be in the reverse order to that found experimentally. Very interesting improvements in the rolling ball technique were described by Peterson (1954 b), and previous conclusions confirmed and extended, the redistribution of charge by surface conduction and gas discharge being gone into in considerable detail. A cleaning procedure was found which guaranteed reproducible results from Ni/silica, but this seems more likely to have been the consequence of standardizing the contamination, than of completely cleaning the surfaces. The measurements were normally made by allowing the charge to pass from the ball to the electrometer by gas discharge in nitrogen, and prolonged rolling was found to alter the charging characteristics. Peterson states that very great changes in surface properties (including reversals of the sign of charging) are produced by even slight contamination. This is, indeed, characteristic of 'triboelectrification', and clean conditions are almost relegated to the status of an impracticable ideal by the very nature of the experiments. The work of Debeau (1944) on the effect of adsorbed gas films demonstrates how important the chemical state of a surface is, in determining how it will charge.

Kunkel (1950) investigated the charge on dust particles after dispersion into a cloud; charges on individual particles were measured by

the Hopper and Laby technique (1941). There were good reasons for believing that the charge separation took place at the first dispersal of the powder aggregate, and was not due to subsequent friction between particles. The work was extended by Dodd (1953), who found no effect of varying humidity on the charging of soft soda-glass. Since, however, the particles had no opportunity of discharging after the first dispersal,

Fig. 12



From *Proc. roy. Soc.*

this is compatible with Peterson's observation of a marked humidity variation, accepting that this was entirely due to a loss of charge originally transferred. Kunkel found that his particles were about equally likely to be positively as negatively charged, the r.m.s. charge increasing with size of particle. He proposed a theory to account for the charging, similar to that of Natanson (§ 2.1) for the statistical fluctuation charging of droplets. The theory depends on the tendency of the particles to

adhere, and applies to the charging of like surfaces when the bridge formed by adhesion is broken. Kunkel points out that there will be lattice imperfections in the bridge which will give rise to localized electron levels, and these may contain electrons drawn from the rest of the lattice. When the bridge is fractured, it may happen that an electron remains on one side of the fracture, but the compensating positive charge on the other, thus separating charge to the extent of one electron. If the fracture passes close to a number of localized levels, it is uncertain which part will get the electrons. There will therefore be a fortuitous charging, of magnitude proportional to the square root of the number of localized levels involved, and of random sign. This mechanism provides a reasonable explanation of Kunkel's experimental results, in particular for quartz dust which was quite well cleaned. It cannot, however be the correct explanation, since Harper (1955) has shown that the charging is independent of adhesion.

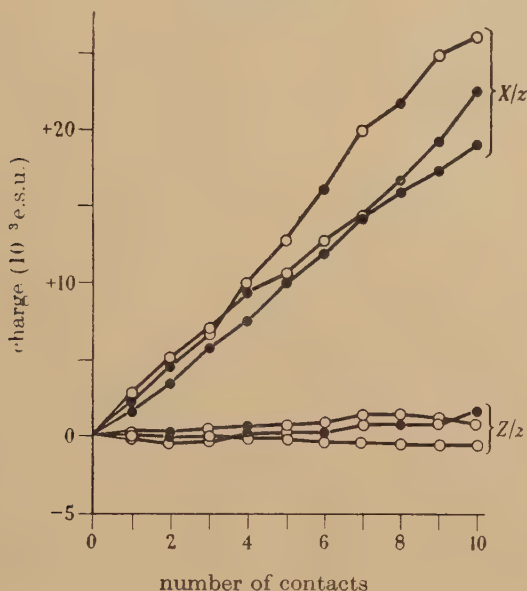
An investigation of the charging of quartz and silica surfaces, in which special attention was paid to cleaning them (including a technique for experimenting with recently fused silica), has been carried out by Harper (1955). It was found that different crystal faces of quartz charged each other as strongly as unlike materials; indeed the charging obtained was at least an order of magnitude larger than that formerly obtained with any materials under any conditions. Such charging must be a true surface effect. An *X* cut face was positive with respect to a *Z* cut, typical sets of charging curves being shown for *X* cut/*Z* cut and *Z* cut/*Z* cut in fig. 13. The charging was not piezoelectric in origin, since it did not reverse when an *X*<sub>-</sub> face was used instead of an *X*<sub>+</sub>. The average charging for *X* cut/*Z* cut corresponded to a chance of 1 in 30 that an electron would be transferred per lattice square, allowing for 'tunnel effect' according to the theory of electronic transfer, or 1 in 15 for positive ionic transfer assuming ions to be transferred only over the area of true contact. These figures are only approximate; even greater probabilities may occur.

If surface electron levels exist on quartz, they may be expected to be different on different crystal faces: if the *X* cut/*Z* cut charging is to be explained by the theory proposed for electronic transfer between solids, a difference of 4 v between the levels concerned is required to account for the observed charging. Now these surface levels must lie between the full and empty bands for quartz. Since the ultra-violet absorption sets in at about 2000 Å, the gap between the full and the empty bands must be at least 6 v, and this is compatible with a difference of 4 v between the surface levels on different crystal faces. It must be realized, however, that, after the transfer of charge, if not before, the electron levels concerned in the transfer can be neither practically full nor practically empty: they must be partially filled. Since they do not give rise to surface conductivity, they can hardly belong to the entire surface of the insulator, so they cannot be Tamm levels (Statz 1950). They must therefore be very localized, and belong either to dislocations, or to contamination.



They might be dislocation levels, though such an explanation would seem to call for devitrified fused silica to be almost a conductor. Contamination, particularly oxygen, can easily supply acceptor levels, but it is very difficult to see how sufficiently shallow donor levels can be present on a quartz surface. The hypothesis of electronic transfer, although it gives the right order of magnitude of charging, nevertheless weakens when subjected to closer inspection.

Fig. 13



*From Proc. roy. Soc.*

Upper curves, charge on X face (against  $z$ ); lower curves, charge on Z face (against  $z$ ). ● sticking; ○ not sticking.

Perhaps, then, electrophilic insulators share with electrophobic insulators an inability to charge by electronic transfer, electrophilic insulators being distinguished by an ability to transfer positive ions. There is a slight clue from the experimental evidence which points in this direction. Before a fully satisfactory technique for cleaning quartz and similar surfaces had been established, the author was much troubled by the occasional complete disappearance of charging. This was almost certainly caused by a film of grease, the surface losing, not only its ability to charge, but also its wettability. Such films can migrate over considerable distances, particularly when the surface is very clean (Bowden and Tabor 1950, p. 77). Under the conditions of the experiment, the films were probably monomolecular. If this was the case, then the film would not have been thick enough to stop charging by electronic transfer: electrons can 'tunnel' through much greater distances. A monolayer, however, might well stop the transfer of positive ions.

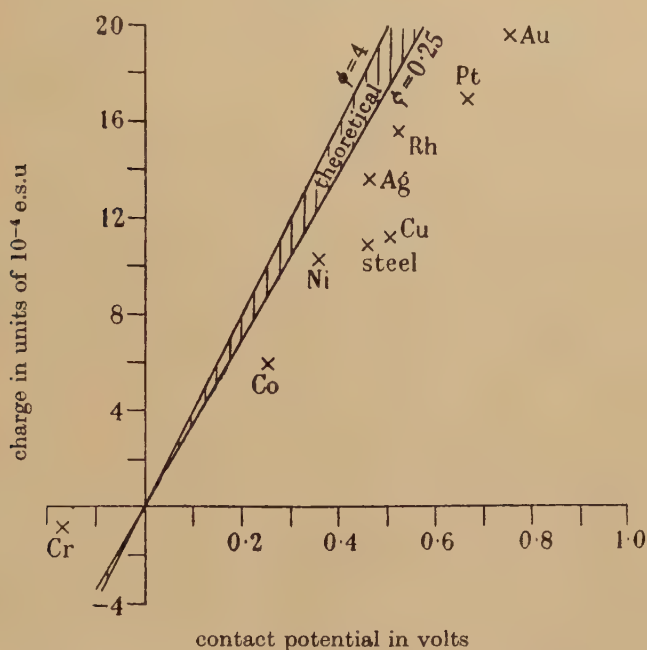
The positive ions need not necessarily come from the crystal lattice, even though the surface is not contaminated in the usual sense of the word. We must beware of assuming that the well-cleaned state is that of the interior structure terminating at the surface in a mathematical discontinuity. Such a state may be chemically unstable, and stabilize itself by forming a structurally different surface layer. The formation of veils (§ 2.3) may be associated with such a process. The emission of 'Exo-electrons' (Kramer 1949), proves that considerable energy changes are involved in the ageing of surfaces, though too few 'Exo-electrons' are emitted to be significant for the mechanism of charge transfer. Electrophilic insulators are hydrophilic, so they might stabilize their surface by attaching a number of hydroxyl radicles, thereby becoming negatively charged, with the relevant electrons deeply bound. The compensating positive ions could be hydrated hydrogen ions, loosely bound, possibly held mainly by image forces. It is not to be expected that different crystal faces would stabilize in the same way; on contact, therefore, there could be a migration of the loosely bound positive ions from one surface to the other (Henry 1953 b), thus giving rise to charging. If the controlling factor was image force, the hypothesis would provide an explanation of Coehn's rule that the surface possessing the higher dielectric constant is the one that becomes positively charged. Unfortunately, however, there are many exceptions to this rule, and it is contradicted by the charging of quartz surfaces, if we take the relevant dielectric constant as being the one perpendicular to the surface. Quartz/quartz charging would seem to be a crucial test; we must conclude that the truth of the matter is more complicated than we have been supposing. The phenomenon of pyroelectricity, however, suggests that we may be near the truth. A pyroelectric crystal, such as tourmaline, fractures with opposing faces oppositely charged (Handbuch der Experimentalphysik, 1930). In the course of time, however, ions from the surroundings neutralize the charges on the surfaces, which were initially polarization charges. If now the crystal is heated in a flame, the polarization changes with increase in temperature, but the new polarization charges are at once compensated by ions from the flame. On removing the crystal from the flame, and allowing it to cool, the polarization reverts to its former value, the ionic charges remain the same, so the crystal appears to develop a charge on cooling, whereas it did not do so on heating. It seems likely that many surfaces, no longer *in statu nascendi*, though otherwise to be regarded as clean, are in fact covered with loosely bound positive (or negative) ions, and that these ions play an essential part in the charging of insulators. Electrophilic insulators would then be those possessing such a surface layer of ions, and electrophobic insulators those without.

#### 5.4. *Metals and Oxides*

Figure 14, from Harper (1951 a), compares the correlation of the 'separation charging' of metals with contact potential, found experimentally, with that predicted by theory. There are no adjustable

constants in the theory. The experimental charging is seen to be quite well predicted by theory, though somewhat too small. The discrepancy is, however, accounted for by the theory of the effect of protuberances (§ 5.2), if these ranged in height between 200 and 9000 Å for different balls. This is very reasonable for balls electroplated with metals which included both those that are difficult to electrodeposit, and those that come down with a perfect mirror finish. The consistently low values of the experimental points is probably to be explained, therefore, as being due to surface imperfections, the theory then accounting quantitatively for the charging of metal/metal interfaces. Strictly, however, the surfaces were not metallic. It is well known (Evans 1946) that nearly all clean metals rapidly acquire an invisible protective film, when exposed to a

Fig. 14



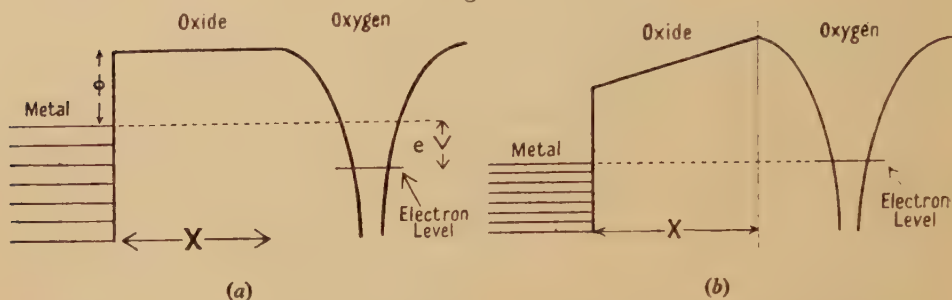
From *Proc. roy. Soc.*

laboratory atmosphere, the film consisting mainly of oxide. A theoretical approach to the formation of such films has been given by Cabrera and Mott (1949). Figure 15 shows two stages in the formation of a thin oxide layer, with an adsorbed layer of oxygen atoms on the oxide, according to their theory. When equilibrium conditions obtain, the electron level in the adsorbed oxygen atoms is partially filled, and so can behave both as a donor and an acceptor level, any change in population being shared with the Fermi level in the metal if the oxide layer is not so thick as to behave like an insulator. The depth of the oxygen level is about  $5\frac{1}{2}$  v, and not greatly dependent on the nature of the metal, and will therefore usually be too deep for it to act as a donor level to an



insulator, though able to do so to another similarly oxidized metal in which the corresponding level is deeper. Invisibly oxidized metal surfaces, therefore, conform to the requirements of the theory of 'separation charging'; the only difference being that the relevant contact potential is not that for clean metals, being in general less. A sufficiently thick oxide layer will behave like a semiconductor (§ 5.2),

Fig. 15



From Rep. Progr. Phys.

Electronic levels in the metal, oxide and adsorbed oxygen : (a) before electrons have passed through the oxide, (b) when equilibrium is set up.

and the thickness at which the change from behaving like a metal to behaving like a semiconductor occurs, will depend on the concentration of 'traps' in the oxide. The author has noticed that invisibly oxidized and obviously tarnished metals usually behave differently, though visible tarnish can, on occasion, behave like invisible tarnish. The visible tarnish is, nearly always, negative with respect to the invisible.

### 5.5. The Significance of Friction

We have seen that the mere contact of two surfaces, followed by their separation, causes them, in general, to become highly charged. Since rubbing includes contact, it is tempting to conclude that Volta was right when he taught that electrification by friction arises from an effect of contact, the virtue of friction being merely to multiply the points of contact. There is not much experimental evidence as to how far this explanation accounts for the phenomena of electrification by rubbing, but it must at least be a correct approach to an understanding of the phenomena, provided allowance is made for changes in the surfaces brought about by rubbing. We noted (§ 5.1) the importance of the precise manner of breaking contact, when discussing the work of Knoblauch (1902). Shaw (1917) noticed that a light rub was more effective than a vigorous one; this suggests that rubbing does more to spoil contact charging than to provide a better mechanism. Alternative mechanisms may, however, prove to be important, particularly with electrophobic insulators. Frenkel (1941) has put forward a theory in which hot spots (Bowden 1936), formed during rubbing, turn the insulator into a semiconductor at the contact points. Henry (1953 a) has shown that, when

two sheets of similar material are rubbed together, each sheet becomes covered with patches of charge of different sign, though neither may possess a resultant charge. The effect was shown up by sprinkling the surfaces with a mixture of coloured powders which charged each other with opposite signs when scattered from the sifter, the powders then being attracted to the oppositely charged areas on the surface under test. Plate I shows the result of such an experiment on polythene. The dark spots were positively charged, and the grey areas negatively, whilst the light areas did not collect any powder and presumably carried little charge of either sign. It is clear that we have to do with an effect of asymmetric rubbing, protuberances behaving differently from flatter parts. Henry suggests that the separation of charge by the asymmetric rubbing of like surfaces is due to a thermal gradient across the surfaces, and is a non-equilibrium effect.

It is well known (Bowden and Tabor 1950) that friction between surfaces frequently leads to a bulk transfer of the material of the one surface to the other, on a very small scale. It seems likely that the particles transferred will be electrically charged, and this would result in a form of 'entrainment' charging between solids. Some remarkable experiments of Kluge (1929) on the rubbing of silk against metals in a high vacuum may have involved this mechanism. It could well be of importance in the case of ordinary surfaces covered with relatively thick layers of dirt easily removed, or very clean surfaces which tend to adhere strongly to each other. Harper (1955), however, found that the adhesion of very clean quartz surfaces did not affect the charging.

## § 6. CONCLUSIONS, AND A WORKING HYPOTHESIS

In reviewing the various situations in which static charge is generated, we have found two only for which we can be quite sure of the mechanism responsible, and these two are unfortunately not intrinsically important. For them, however, detailed mathematical theories have been shown to give a quantitative account of the experimental findings. One case is the light contact of invisibly tarnished metals in which the charging is due to electronic transfer establishing the contact potential by 'tunneling'; the other is the entrainment charging of medium sized particles of liquids of very low conductivity, in which the mechanical disintegration of the liquid transfers electrolytic ions, already present in the liquid, to the gas. It is also fairly certain that charging by freezing results from the preferential freezing in of electrolytic ions, and that the charging of supposedly insulating liquids by passing them through tubes is the result of preferential adsorption of electrolytic ions of one sign on the walls of the tube, the other sign being carried on. A mathematical theory of this process is available, but we found reason to revise Cooper's rule for safe working when handling dangerous liquids (§ 3.1). The charging of a gas when passing through an orifice, on the other hand, only occurs when particulate matter is present.

There seems to be no reason to doubt that, when large ions are formed during bubbling, splashing, spraying or the shattering of drops, they come from the evaporation of the smallest particles formed by mechanical disintegration from the thin sheets of liquid which arise in these processes. The orthodox explanation of the negative ions obtained from water in this way is that a double layer of ions at the surface is split, one half supplying the large ions. Careful consideration of the nature of a water surface, however, and how it should behave during entrainment, shows the explanation to be tenable only if unwanted ionizable capillary active contamination was present in the experiments. An alternative explanation depends on a redistribution of statistical fluctuations of charge during the formation of satellite droplets alongside larger drops during the disintegration of the liquid film. There is also the possibility of large ions being formed by the entrainment of particulate matter on which electrolytic ions were adsorbed in the liquid. Such particulate matter is widespread, and extremely difficult to get rid of. It occurs as motes in liquids, and the veils on crystals, which are left behind after dissolving the crystal, retaining the shape of the crystal.

The division of opinion as to whether pure mercury gives positive ions by entrainment, or none at all, is resolved by the theoretical consideration that the ions, if any, should be negative. The positive ions found experimentally were probably due to the presence of water or oxide, oxygen being of critical importance in certain cases of mercury insulator charging.

Mercury/insulator charging is important in providing intimacy of contact without friction; the large charges obtainable indicate that the Volta-Helmholtz hypothesis that 'frictional' charging is really contact charging, is capable of providing ample charge. It is probably the most effective mechanism normally operating, but specifically frictional effects can seriously alter the charging, and may be fundamental in the asymmetric rubbing of like surfaces. There is little evidence concerning the importance of bulk transfer from one surface to the other.

It is most unlikely that the charging of mercury/glass can be accounted for by the transfer of electrons, and, whenever one tries to account for charging in which insulators take part, the assumption of electronic transfer leads to difficulties. It seems, in fact, that, if insulators take part, electrons do not, and that electronic transfer is confined to metals and semiconductors, though the latter cannot be expected to acquire much charge in this way. The existence of charging by ionic transfer is well established, but it is unlikely that the hypothesis of the division of a double layer of electrolytic ions in an adsorbed water layer is often applicable; charging can still occur at zero relative humidity; high relative humidity may facilitate discharging.

When speaking of static electrification, one is most often thinking of the case in which one surface at least is a good insulator. When clean, good insulators fall into two classes. The electrophobic ones do not



charge at all by light contact. Very remarkably, amber, after which electricity is named, is an electrophobic insulator. Electrophilic insulators charge freely, and the largest charging so far observed was obtained with two different crystal faces of quartz, which demonstrates, for this case at least, that the charging was a truly surface phenomenon. Unless very special precautions are taken, electrophilic surfaces are liable to be contaminated with electrophobic dirt (e.g. grease), and this reduces the charging; electrophilic contamination on an electrophobic substrate can give rise to charging. Thus, in practice, charging seems always to be possible, though never reaching the high values attained with really clean electrophilic surfaces.

If it be true that the generation of static charge on insulators is mainly an effect of contact in which positive ions move, we have to account for the availability of the positive ions on electrophilic insulators. We may suppose, as a working hypothesis, that the termination of the interior structure of these insulators as a mathematical discontinuity at the surface, as would be the immediate result of fracture, is chemically unstable, and that the first step towards stabilization under normal conditions consists in the attachment of hydroxyl radicals. Since, however, this would give the surface a negative charge, a second step would be the neutralization of this charge by the attachment of positive ions from the surroundings, possibly hydrated hydrogen ions. These ions would be loosely bound, and therefore be easily transferred to another surface.

#### ACKNOWLEDGMENTS

In preparing this review, the author has changed his mind about the interpretation of many of the experimental facts. This is partly the consequence of a reconsideration of the published literature, but even more the result of stimulating suggestions made by Professor N. F. Mott, and of discussions with Professor M. Blackman. To them he is much indebted. He is also indebted to the University Library, Heidelberg, for making possible the publication of the photograph from Hochschwender's dissertation. Dr. N. Hast kindly supplied the electron micrographs of veils.

#### REFERENCES

- AGANIN, M., 1914, *Ann. Phys., Lpz.*, **45**, 1013.  
 ALTY, T., 1924, *Proc. roy. Soc. A*, **106**, 315.  
 ARTHUR, D. F., 1955, *J. Text. Inst.*, **46**, 1721.  
 BATEMAN, H., 1911, *Phil. Mag.*, **21**, 745.  
 BECKER, A., 1909, *Ann. Phys., Lpz.*, **29**, 909.  
 BLANCHARD, D. C., 1954, *Nature, Lond.*, **173**, 1048.  
 BLOCH, L., 1911, *Ann. Phys., Paris*, **23**, 28.  
 BOWDEN, F. P., 1936, *Proc. roy. Soc. A*, **154**, 640.  
 BOWDEN, F. P., and TABOR, D., 1950, *The friction and lubrication of solids* (Oxford: University Press).  
 BÜHL, A., 1926, *Ann. Phys., Lpz.*, **80**, 137.  
 BUSSE, W., 1925, *Ann. Phys., Lpz.*, **76**, 493.

- BUTLER, J. A. V., 1951, *Electrical phenomena at interfaces* (London : Methuen and Co.).
- CABRERA, N., and MOTT, N. F., 1949, *Rep. Progr. Phys.*, **12**, 163.
- CHALMERS, J. A., and PASQUILL, F., 1937, *Phil. Mag.*, **23**, 88.
- CHAPMAN, S., 1934, *Physics*, **5**, 150 ; 1937, *Phys. Rev.*, **52**, 184 ; 1938, *Ibid.*, **54**, 520, 528.
- CHRISTIANSEN, C., 1894, *Ann. Phys., Lpz.*, **53**, 408.
- COEHN, A., and MOZER, H., 1914, *Ann. Phys., Lpz.*, **43**, 1048.
- COEHN, A., and LOTZ, A., 1921, *Z. Phys.*, **5**, 242.
- COEHN, A., and CURS, A., 1924, *Z. Phys.*, **29**, 186.
- COEHN, A., and DUHME, E., 1924, *Z. Phys.*, **27**, 358.
- COOKE, B. A., 1955, *Nature, Lond.*, **176**, 264.
- COOPER, W. F., 1953, *Brit. J. appl. Phys.*, Supp. 2, 11.
- CURRIE, B. W., and ALTY, T., 1929, *Proc. roy. Soc. A*, **122**, 622.
- DAVIES, J. T., and RIDEAL, Sir E., 1955, *Canad. J. Chem.*, **33**, 947.
- DEAGLIO, R., 1928, *Z. Phys.*, **51**, 279.
- DEBEAU, D. E., 1944, *Phys. Rev.*, **66**, 9.
- DE BROGLIE, M., 1909, *Ann. Phys., Paris*, **16**, 5 ; 1910, *C.R. Acad. Sci., Paris*, **150**, 1113.
- DODD, E. E., 1953, *J. appl. Phys.*, **24**, 73.
- DOMBROWSKI, N., and FRASER, R. P., 1954, *Phil. Trans. A*, **247**, 101.
- DORSEY, N. E., 1940, *Properties of ordinary water substance* (New York : Reinhold).
- EHRENBERG, W., and HÖNL, H., 1931, *Z. Phys.*, **68**, 289.
- ELSTER, J., and GEITEL, H., 1890, *Wiener Ber.* 99.
- EVANS, U. R., 1946, *Metallic corrosion, passivity and protection* (London : Arnold and Co.).
- EVE, A. S., 1907, *Phil. Mag.*, **14**, 382.
- FOWLER, R. H., 1933, *Proc. roy. Soc. A*, **141**, 56.
- FRENKEL, J., 1941, *J. Phys. U.S.S.R.*, **5**, 25.
- FREUNDLICH, H., 1926, *Colloid and Capillary Chemistry*, 3rd ed., 284 (London : Methuen and Co.).
- FRUMKIN, A., and OBRUTCHEWA, A., 1931, *Kolloid Z.*, **54**, 2.
- GILL, E. W. B., 1953 a, *Brit. J. appl. Phys.*, Supp. 2, 22 ; 1953 b, *Ibid.*, Supp. 2, 16 ; 1954, *Nature, Lond.*, **173**, 398.
- GILL, E. W. B., and ALFREY, G. F., 1949 a, *Nature, Lond.*, **164**, 1003 ; 1949 b, *Ibid.*, **163**, 172 ; 1952 a, *Proc. phys. Soc. Lond. B*, **65**, 546 ; 1952 b, *Nature, Lond.*, **169**, 204.
- GURNEY, R. W., 1936, *Ions in solution* (Cambridge : University Press).
- HAMPEL, B., and LUTHER, H., 1955, *Naturwiss.*, **42**, 252.
- Handbuch der Experimentalphysik*, 1930, **10**, 327 (Leipzig : Akademische Verlagsgesellschaft m.b.H.).
- HARPER, W. R., 1951 a, *Proc. roy. Soc. A*, **205**, 83 ; 1951 b, *Nature, Lond.*, **167**, 400 ; 1953 a, *Proc. roy. Soc. A*, **218**, 111 ; 1953 b, *Brit. J. appl. Phys.*, Supp. 2, 19 ; 1955, *Proc. roy. Soc. A*, **231**, 388.
- HAST, N., 1952, *Arkiv för Fysik*, **4**, 37.
- HENRY, P. S. H., 1953 a, *Brit. J. appl. Phys.*, Supp. 2, 31 ; 1953 b, *Ibid.*, Supp. 2, 6.
- HOCHSCHWENDER, E., 1919, *Diss* (Heidelberg).
- HOPPER, V. D., and LABY, T. H., 1941, *Proc. roy. Soc. A*, **178**, 243.
- JOHONNOT, E. S., 1899, *Phil. Mag.*, **47**, 501.
- KELLNER, L., 1952, *Rep. Progr. Phys.*, **15**, 1.
- KELVIN, Lord, 1867, see Thomson, Sir W., *Proc. roy. Soc.*, **16**, 67.
- KLUGE, W., 1929, *Ann. Phys., Lpz.*, **1**, 1.
- KNOBLAUCH, O., 1902, *Z. phys. Chem.*, **39**, 225.

- KRAMER, J., 1949, *Z. Phys.*, **125**, 739.
- KUNKEL, W. B., 1950, *J. appl. Phys.*, **21**, 820.
- LATIMER, W. M., and RODEBUSH, W. H., 1920, *J. Amer. Chem. Soc.*, **42**, 1419.
- LENARD, P., 1892, *Ann. Phys., Lpz.*, **46**, 584; 1915, *Ibid.*, **47**, 463; 1921, *Ibid.*, **65**, 629.
- LOEB, L. B., 1945, *Science*, **102**, 573.
- MASON, B. J., 1953 a, *Quart. J. R. met. Soc.*, **79**, 501; 1953 b, *Tellus*, **5**, 446; 1954, *Nature, Lond.*, **174**, 470.
- MEDLEY, J. A., 1950, *Nature, Lond.*, **166**, 524; 1953 a, *Brit. J. appl. Phys.*, Supp. 2, 28; 1953 b, *Nature, Lond.*, **171**, 1077; 1953 c, *Brit. J. appl. Phys.*, Supp. 2, 23.
- MORANT, M. J., 1954, *J. appl. Phys.*, **25**, 1053.
- MOTT, N. F., 1938, *Proc. Camb. phil. Soc.*, **34**, 568.
- MOTT, N. F., and GURNEY, R. W., 1940, *Electronic processes in ionic crystals* (Oxford: University Press).
- NATANSON, G. L., 1949, *Zhur. Fiz. Khim.*, **23**, 304.
- NEWITT, D. M., DOMBROWSKI, N., and KNELMAN, F. H., 1954, *Trans. Instn chem. Engrs, Lond.*, **32**, 244.
- NOLAN, J. J., 1914, *Proc. roy. Soc. A*, **90**, 531.
- OWSTON, P. G., 1951, *Chem. Rev.*, **5**, 4.
- PERUCCA, E., 1921 a, *N. Cimento*, **21**, 34; 1921 b, *Ibid.*, **21**, 275; 1921 c, *Ibid.*, **22**, 56.
- PETERSON, J. W., 1954 a, *J. appl. Phys.*, **25**, 501; 1954 b, *Ibid.*, **25**, 907.
- POMEROY, J. C., 1908, *Phys. Rev.*, **27**, 492.
- RAMAN, C. V., and RAMDAS, L. A., 1927, *Phil. Mag.*, **3**, 220.
- RUDGE, W. A. D., 1912, *Phil. Mag.*, **23**, 852; 1913, *Ibid.*, **25**, 481; 1914, *Proc. roy. Soc. A*, **90**, 256.
- RUSSELL, A., 1909, *Proc. roy. Soc. A*, **82**, 524.
- SHAW, P. E., 1917, *Proc. roy. Soc. A*, **94**, 16.
- SIMPSON, G. C., 1909, *Phil. Trans.*, **209**, 379.
- SKINNER, S. M., 1955, *J. appl. Phys.*, **26**, 498.
- SKINNER, S. M., SAVAGE, R. H., and RUTZLER, J. E., 1953, *J. appl. Phys.*, **24**, 438; 1954, *Ibid.*, **25**, 1055.
- SOMMERFELD, A., and BETHE, H., 1933, *Handbuch der Physik*, 2nd ed. 24/2, 441 (Berlin: Springer).
- STATZ, H., 1950, *Z. Naturf.* **5a**, 534.
- VICK, F. A., 1953, *Brit. J. appl. Phys.*, Supp. 2, 1.
- WESENDONCK, K., 1892, *Ann. Phys., Lpz.*, **47**, 529.
- WEST, D. C., 1954, *J. appl. Phys.*, **25**, 1054.
- WOODCOCK, A. H., KÜNTZLER, C. F., ARONS, A. B., and BLANCHARD, D. C., 1953, *Nature, Lond.*, **172**, 1144.
- WORKMAN, E. J., and REYNOLDS, S. E., 1950, *Phys. Rev.*, **78**, 254; 1952, *Nature, Lond.*, **169**, 1108.
- WORTHINGTON, A. M., 1895, *The splash of a drop* (London: Society for Promoting Christian Knowledge).
- ZELENY, J., 1933, *Phys. Rev.*, **44**, 837.



# A Theory of the Fracture of Metals

By A. N. STROH

Department of Physics, University of Sheffield

## CONTENTS

§ 1. INTRODUCTION.	
§ 2. PRELIMINARY CONSIDERATIONS.	
2.1. The strength of the ideal lattice.	
2.2. Griffith's theory.	
§ 3. THE BRITTLE FRACTURE OF METALS.	
3.1. Piled up groups of dislocations.	
3.2. The production of a crack.	
3.3. The growth of the crack.	
3.4. Comparison with experimental results.	
3.5. Hydrogen embrittlement.	
3.6. Slip near a piled up group.	
§ 4. THE LOWER YIELD POINT.	
4.1. The stress needed to move a dislocation.	
§ 5. THE BRITTLE-DUCTILE TRANSITION.	
5.1. The probability of brittle fracture.	
5.2. Effect of strain rate.	
5.3. Effect of grain size.	
5.4. The activation energy.	
5.5. Notch brittleness.	
§ 6. THE PROPAGATION OF CRACKS.	
6.1. Griffith's theory applied to metals.	
6.2. The velocity of a crack.	
6.3. The propagation transition.	
6.4. The effect of transverse stress.	
§ 7. DUCTILE FRACTURE.	
ACKNOWLEDGMENTS.	
APPENDIX A : The energy of a crack.	
B : The initial growth of a crack.	
REFERENCES	

## § 1. INTRODUCTION

THE topics which will be considered here can broadly be described as the transcrystalline fracture of polycrystalline materials ; an attempt will be made to show how the principal features observed follow from a simple model. Single crystals will not be considered in detail ; though they have been the subject of some recent experimental studies, e.g. Deruyttère and Greenough (1955), Allen *et al.* (1955), it has not yet proved possible to give a satisfactory account of their properties. Other forms of fracture, for example, fatigue, fracture under creep conditions, will also be excluded, even though they may be of great technical

importance and the subject of much experimental work, for it seems likely that quite different mechanisms operate in such circumstances. A recent paper by Mott (1956) includes a discussion of fatigue in addition to those topics now considered while a review by Petch (1954) deals with the whole problem of fracture and includes a more complete account of the experimental results.

It should be mentioned that other theories of fracture, such as those of Fisher (1955) and Kochendörfer (1954) which are quite different from that presented now, have been proposed; such treatments will not be discussed here as they have yet to be developed in sufficient detail to permit of a quantitative comparison with experimental results.

Two main types of fracture fall within our consideration. First the metal may be brittle and break with only little plastic deformation; after fracture, the surfaces show bright crystalline facets, on which cleavage has occurred. Here definite crystalline planes are involved, for example, body-centred cubic iron cleaves on  $\{001\}$  planes, and zinc (hexagonal) on the basal plane. No metal with a face-centred cubic structure is known to break in this way, and other metals do so only at low temperatures. Such fractures spread with great rapidity, and in practice may produce spectacular failures. They attracted particular notice during the last war in connection with welded ships; some 200 Liberty ships developed serious cracks, while about a dozen broke completely in two.

The other possibility is that the metal is ductile and plastic deformation precedes fracture. In a tensile test the specimen first necks and later a crack develops in the centre of the neck; after the specimen has grown to sufficient size, the mechanism changes and the final separation is by shear at an angle of  $45^\circ$  to the stress axis; thus the familiar 'cup and cone' is produced. In contrast to the brittle case, the fracture surface now has a matt or fibrous appearance. Ductile cracks spread only slowly and can be stopped by a decrease in the load.

For the very reason that complications due to plastic flow are absent, brittle fracture is the simpler to interpret, and we shall find it possible to give a much more complete account of it than we can of ductile fracture. The temperature at which the transition from ductile to brittle behaviour occurs is a matter of some importance for which our model will be able to account. Finally it is found that, with a material such as iron having a sharp yield point, the lower yield point, that is the stress at which a band of deformation (a Lüders band) spreads through the specimen, has a number of features in common with the brittle strength; it is therefore convenient to include a discussion of this topic.

## § 2. PRELIMINARY CONSIDERATIONS

### 2.1. *The Strength of the Ideal Lattice*

It has been known for a long time that a substance having a perfect crystalline structure should exhibit a strength very much greater than is





## 2.2. Griffith's Theory

To account for the discrepancy between the observed and calculated values of the strength, Griffith (1920) assumed that a real material contains a number of small cracks which can act as stress concentrators raising the stress at their tips to the ideal strength  $\sigma_m$ ; in this way the crack can extend and produce fracture of the material.

Consider the two dimensional (plane strain) problem of a crack of length  $c$  extending through a thick plate, and subjected to a tensile stress  $\sigma$  normal to the crack. From Inglis's calculation (1913) of the stress distribution, the elastic energy is decreased due to the presence of the crack by an amount  $\pi(1-\nu)\sigma^2c^2/8G$ , where  $G$  is the rigidity modulus and  $\nu$  Poisson's ratio. In addition the crack has surface energy  $2c\gamma$ , so that the total energy of the crack is

$$W = -\pi(1-\nu)\sigma^2c^2/8G + 2c\gamma. \quad (2)$$

If the crack is to extend,  $W$  must decrease as  $c$  increases; the equilibrium length of the crack is given by  $dW/dc=0$ , and from eqn. (2) equilibrium will occur when

$$\sigma = \{8\gamma G/\pi(1-\nu)c\}^{1/2}. \quad (3)$$

The crack can extend if this stress is exceeded, and so eqn. (3) gives the fracture stress of the material. There is clearly always a crack length  $c$  which will lead to the observed strength.

Sack (1946) has extended the calculation to a circular (penny-shaped) crack; the only change is that the critical stress is greater by a factor  $\frac{1}{2}\pi (=1.57)$  than is given by eqn. (3). This result is interesting because most calculations on the behaviour of cracks have, for simplicity, dealt with a two-dimensional system only; Sack's work suggests that no very great error will be made thereby.

An alternative approach is to use the condition that, for the crack to extend, the stress at its tip must have the value  $\sigma_m$  (given by eqn. (1)); over an area of atomic dimensions. By considering the stress magnification produced by the crack, Orowan (1949, 1955) has shown that this leads, within the accuracy of the calculation, to the same critical stress as before (eqn. (3)). There is thus general agreement as to the stress needed to propagate a crack through a brittle body.

Though the Griffith theory may possibly give a satisfactory account of the fracture of a completely brittle solid, such as glass on which Griffith's own experiments were done, it seems unlikely that pre-existing cracks are important in metals. To account for the observed strength in this way the cracks would often have to be unreasonably large—several millimetres in length, for example, for a weak single crystal of zinc. Such cracks could hardly always be present and escape detection. Nevertheless Griffith's ideas occupy a fundamental position in the theory of fracture and have greatly influenced much of the subsequent work. In particular, an application of Griffith's theory to the propagation of brittle cracks in metals will be discussed in § 6.1.

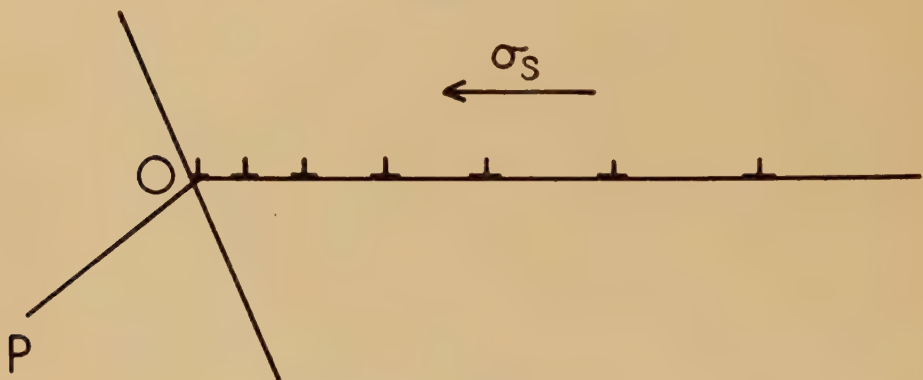
## § 3. THE BRITTLE FRACTURE OF METALS

3.1. *Piled Up Groups of Dislocations*

The presence of dislocations in metal crystals is now well established, and considerable progress has been made in elucidating the part they play in plastic deformation and work hardening; full accounts of their properties are given in the books by Cottrell (1953), Read (1953) and Friedel (1956). It may be expected that dislocations play an equally important role in fracture. Here the edge dislocation, at which a half plane of atoms terminates, is the more important type of dislocation, since it is surrounded by regions of tensile (and compressive) stresses. If a large number of such dislocations arrange themselves with their stress fields reinforcing one another, we may expect that they will jointly be able to produce a stress great enough to cause fracture.

Suppose that a number of dislocations are generated in the same slip plane, and that these spread out until they meet a grain boundary. On reaching the boundary they will be unable to advance further, as in general the slip planes of the next grain will not be suitably orientated. The dislocations will, therefore, be piled up on the grain boundary by the applied stress, and will take up the arrangement shown in fig. 2;

Fig. 2



Piled up group of dislocations.

those at the head of the array will be crowded close together by those behind them, the spacing gradually increasing towards the rear. Such piled up groups, which were first considered by Cottrell (1949) and Eshelby *et al.* (1951), have been observed by Jacquet (1954), and fig. 3, Pl. 14, which is reproduced from one of his photographs, shows a row of etch pits each formed on one dislocation of a piled up group; the whole arrangement agrees so closely with that predicted (a detailed comparison has been made by Bilby and Entwistle 1956) that there can be little doubt that this interpretation is correct.

Near a piled up group the stresses will be large: their nature is most easily seen from Zener's remark (1948) that such a blocked slip line will resemble a freely slipping crack under a *shear* stress<sup>†</sup>; for the fact that the slip line contains a large number of dislocations implies that the material on one side of it can readily slip relative to that on the other. We have, therefore, found an arrangement of dislocations which is very similar to one of Griffith's cracks, and which may be expected to serve equally well in producing fracture.

### 3.2. The Production of a Crack

Let us now consider in what circumstances the stresses round a piled-up group of dislocations will be sufficient to generate a crack. The value of the stresses may be found either by using the analogy mentioned above between a slip line and a crack, or by directly summing the stresses due to the individual dislocations (Stroh 1954). The result is that at a distance  $r$  and from the head of the piled up group the normal stress across the plane OP (fig. 2) is

$$\sigma = \sigma_s (L/r)^{1/2} f(\theta), \quad (r \ll L), \quad . . . . . (4)$$

where the dislocations occupy a length  $L$  in the slip plane,  $\sigma_s$  is the resolved shear stress on the slip plane, and  $f(\theta)$  is a factor depending on the orientation of OP. In a polycrystalline material there will be grains in all orientations and those most favourably situated will break first; hence we may expect fracture to start on a cleavage plane for which  $f(\theta)$  is near its maximum value; in the case of an elastically isotropic material this will be when  $\theta = 70.5^\circ$ . Then if a crack starts at O and extends a distance  $c$  along OP the average value of the stress (4) on it may be written

$$\sigma = \alpha (L/c)^{1/2} \sigma_s, \quad . . . . . (5)$$

where  $\alpha$  is a numerical constant. Inserting this stress in eqn. (2) we find that the production of a crack will result in a decrease of the energy if

$$\sigma_s^2 = 16\gamma G/\pi(1-\nu)\alpha^2 L, \quad . . . . . (6)$$

and this is then the condition for initiating a crack. In Appendix A the constant  $\alpha$  is evaluated and from the results obtained there eqn. (6) may be written

$$\sigma_s^2 = 3\pi\gamma G/8(1-\nu)L. \quad . . . . . (7)$$

Equations (6) and (7) do not contain the length  $c$  of the crack, and so if they are satisfied for one length of the crack, they will be satisfied for all lengths; thus the crack can grow to a finite size with a decrease of energy at every stage. This growth can continue until the length of the crack is comparable to the length of the slip line; then since the stress is no

<sup>†</sup> A crack and a slip line will, of course, behave quite differently under a *normal* stress, since such a stress can act across the slip plane just as if the dislocations were absent; in comparing a slip line with a crack we must therefore disregard any normal component of the stress.





ductility is very slight, there is some evidence to show that fracture is always preceded by a small amount of plastic flow, as the model requires. Low (1954) has found that fracture will not occur until the yield stress has been exceeded. Earlier work has been reviewed by Zener (1948) who emphasizes that a large variety of experiments show that any process which increases the yield stress will also increase the fracture stress, so that initial yielding always occurs before fracture.

### 3.3. The Growth of the Crack

There is an alternative and possibly more vivid way of picturing the origin of a crack, due to Zener (1948) and Stroh (1955a). An edge dislocation represents a half plane of atoms wedged between two complete planes. Now the leading dislocations of a piled-up group are forced very close together; two dislocations will be as close as possible when the extra half plane of atoms associated with each are adjacent, and in this case there is a wedge two atomic spacings thick between the nearest complete planes of atoms. By forcing more dislocations together a wider wedge can be produced until ultimately the neighbouring planes are split apart, so that a crack is formed (fig. 5). Once the crack has started the stresses round it will be relieved and so make it easier to bring in more dislocations, which in turn will widen the wedge and cause the crack to spread further.

Fig. 5



A dislocation with Burgers vector  $3b$ , showing an incipient crack.

On this picture the stress needed to start a crack is the stress which will bring the two leading dislocations of the piled up group within one atomic spacing of each other. Now according to Eshelby *et al.* (1951), the distance between the two leading dislocations is  $3.67 Gb/4\pi(1-\nu)\sigma_s n$ , and equating this to  $b$ , the condition becomes

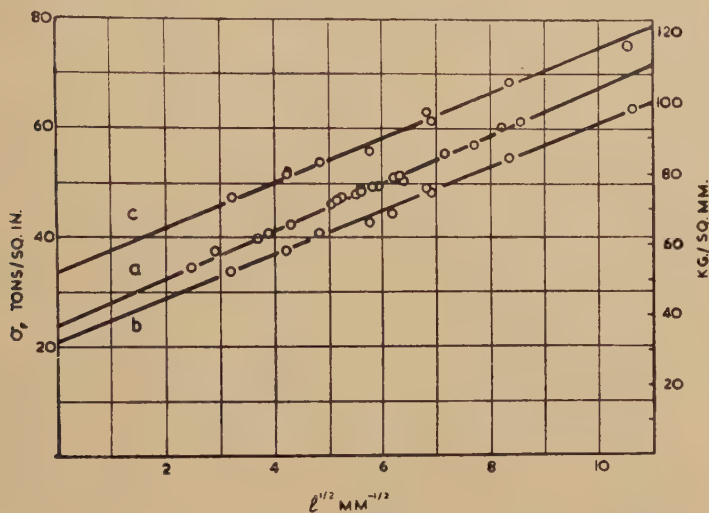
$$nb\sigma_s = 3.67 Gb/4\pi(1-\nu).$$





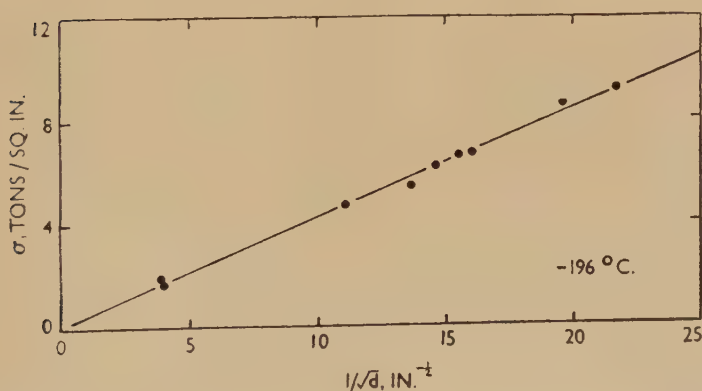
would have been obtained if in eqn. (6) we had replaced  $\sigma_s$  by  $\sigma_s - \sigma_{s0}$ , where  $\sigma_{s0}$  is a constant shear stress; that is if we had assumed that only part of the resolved shear stress was effective in piling up the dislocations. The stress  $\sigma_{s0}$ , which will be discussed further in § 4.1, may be regarded as a frictional force opposing the movement of the dislocations.

Fig. 6



The relationship between fracture strength and grain size for a low carbon steel at (a)  $-196^{\circ}\text{C}$  (brittle), (b)  $18^{\circ}\text{C}$  (ductile), (c)  $-115^{\circ}\text{C}$  (ductile). Petch (1956 a).

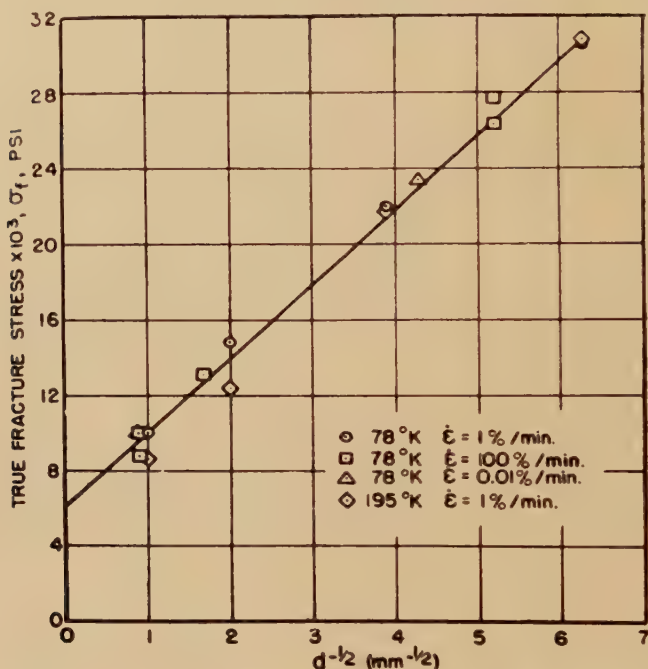
Fig. 7



The relationship between fracture strength and grain size for zinc. Greenwood and Quarrel (1954).

The experimentally determined values of the slope  $K$  of the line  $\sigma_f$  vs  $d^{-1/2}$  may be compared with those calculated<sup>‡</sup> from eqn. (11); as it is seen from table 1, the agreement is very satisfactory.

Fig. 8



The relationship between fracture strength and grain size for magnesium. Hauser *et al.* (1956).

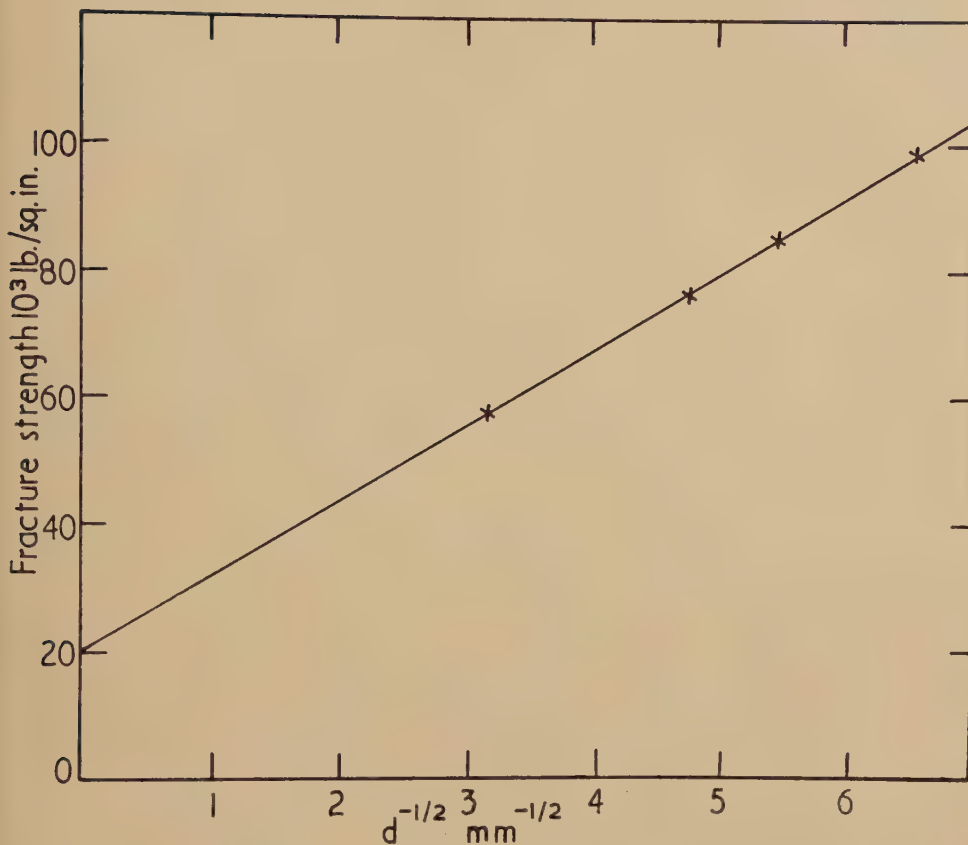
Table 1. Observed and Calculated Values of constant  $K$ 

Metal	Surface energy $\gamma$ erg/cm <sup>2</sup>	Calculated according to (11) c.g.s. units	Observed from graphs c.g.s. units
Fe	1600	$1.9 \times 10^8$	$2.1 \times 10^8$
Zn	800	0.79	1.05
Mg	570	0.52	0.85
Mo	2240 <sup>†</sup>	2.5	2.6

<sup>†</sup> No direct determination of  $\gamma$  for molybdenum is available, and the value used is that estimated by Taylor (1954).

<sup>‡</sup> In calculating the values shown, the elastic constants used are those for the polycrystalline aggregate rather than for a single crystal for we are concerned with a stress field extending across two grains of different orientations, and so want some sort of mean between the elastic constants of the individual grains.

Fig. 9



The relationship between fracture strength and grain size for molybdenum. Bechtold (1953, 1954).

### 3.5. Hydrogen Embrittlement

The importance of the surface energy in determining the fracture strength is seen very clearly in the phenomenon of hydrogen embrittlement. The presence of hydrogen in steel causes cleavage and intercrystalline fracture even at room temperatures, and Petch and Stables (1952) have attributed this to adsorption on the newly formed fracture surfaces with consequent lowering of the effective surface energy. Since the hydrogen will take time to diffuse to the crack this will grow only slowly at first, though once it has grown sufficiently (viz. to the Griffith length corresponding to a clean surface) sudden fracture can occur. Thus there will be a delay time in developing the embrittlement and so the effect will be apparent only for sufficiently long durations of loading; this prediction is confirmed by the results of Petch and Stables shown in table 2. Moreover, diffusion can take place more easily along the grain boundaries than through the interior of the grains so that the



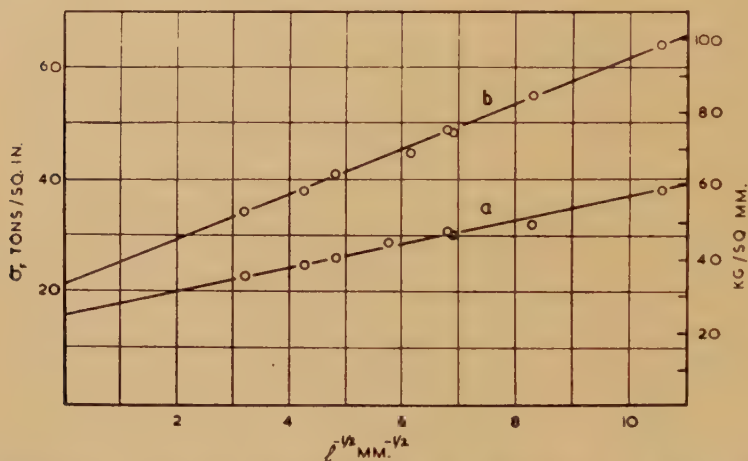
concentration of hydrogen will be greatest near the grain boundaries ; this will favour intercrystalline fracture as is in fact observed.

Table 2. The Effect of Duration of Loading on the Hydrogen Embrittlement of Steel as Measured by the Reduction of Area per cent. Petch and Stables (1952)

Condition	10 <sup>-3</sup> sec.	1-2 min.
Hydrogen free	60.5	61.0
Hydrogen charged	60.5	27.8

If the whole of the effect of the hydrogen can be attributed to a change in the effective surface energy, we should obtain the same dependence of fracture strength on grain size as before (eqn. (12)) but with a reduced value of  $K$  ; that this is so, is seen from the results of Petch (1956 b) shown in fig. 10.

Fig. 10



The relationship between fracture stress and grain size of a mild steel at 18°C (a) with 10 cm<sup>3</sup> hydrogen/100 g, (b) without hydrogen. Petch (1956 b).

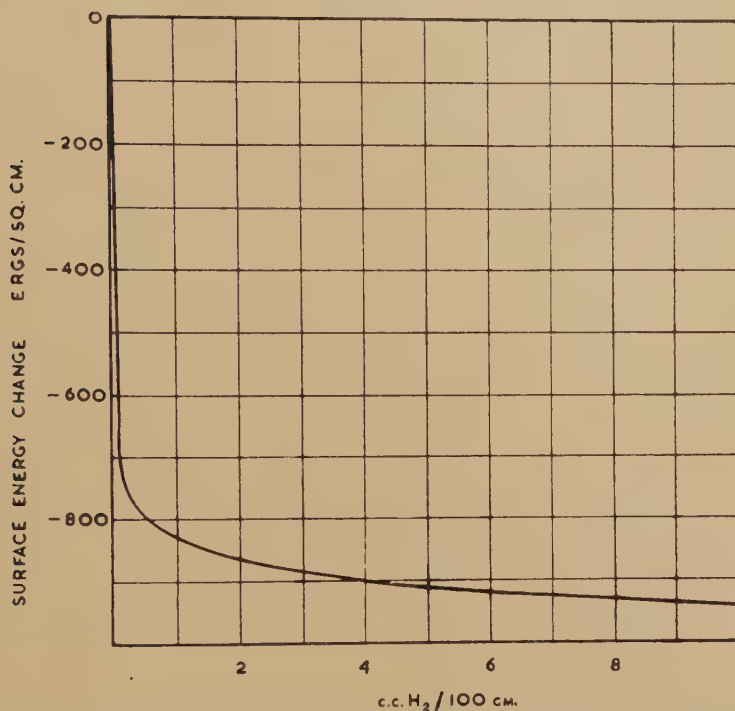
If the adsorption takes place at temperature  $T$  from a diatomic gas at a pressure  $p$ , the gas dissociating on adsorption, Langmuir's isotherm leads to the effective surface energy

$$\gamma_1 = \gamma - 2\Gamma_s kT \log \{1 + (p/p_{1/2})^{1/2}\},$$

where  $\Gamma_s$  is the number of molecules adsorbed per unit area at saturation, and  $p_{1/2}$  is the pressure at which the number of molecules adsorbed is  $\frac{1}{2}\Gamma_s$ . Values of the surface energy calculated in this way for iron with adsorbed hydrogen are shown in fig. 11 ; the effect is large. By substituting  $\gamma_1$  for  $\gamma$  in eqn. (11) the change of strength produced by the

hydrogen may be calculated; for a grain size corresponding to  $d^{-1/2}=5 \text{ mm}^{-1/2}$ , the calculated value of  $\sigma_f - \sigma_0$  is 9.5 tons/in.<sup>2</sup>, agreeing well with the observed value of 11 tons/in.<sup>2</sup>.

Fig. 11



Lowering of surface energy of iron by hydrogen adsorption. Petch (1956 b).

### 3.6. Slip near a Piled Up Group

If a crack is to form near a piled up group of dislocations in the manner we are considering, then it is necessary that the material round the group should remain unyielding; for plastic flow will tend to relieve the stresses here so that they will be insufficient to initiate a crack. Now plastic flow could conceivably start in either of two distinct ways: existing Frank-Read sources nearby might generate new dislocations, or alternatively the stresses might be sufficient to initiate slip in the perfect lattice. The latter possibility will be discussed first.

Frank (1950) has shown that the initiation of slip in the perfect lattice should occur when a shear stress of order  $G/20$  acts over a region whose dimensions are at least 5 to 10 atomic spacings. It is difficult to say from purely theoretical considerations whether such slip will occur more easily than fracture. The problem has been considered by Stroh (1955 a), who concludes that it is not unlikely that fracture will occur first, but no completely satisfactory treatment exists. In both slip

and fracture, the same forces, namely the forces binding the atoms together in the lattice, must be overcome, and so the two processes may be expected to be of comparable difficulty; but present calculations are not sufficiently precise to decide unambiguously which should occur first. We may, however, ask whether the production of new dislocation loops in the perfect lattice helps to explain the observed behaviour of metals; the view taken here is that it does not. For, if slip could be initiated in this way, this would cause the deformation to spread through the material from one grain to the next; it would be natural to identify this process with the spread of a Lüders band. But the stress at which this occurs (the lower yield point) is strongly temperature-dependent, while the energy of a dislocation loop is so large that temperature is unlikely to play any appreciable part in its production: it is therefore difficult to account for the spread of a Lüders band in this way. On the other hand, as will be seen in § 4, the correct temperature-dependence is obtained if slip is always supposed to start from Frank-Read sources already present in the material. We shall therefore reject the possibility that the stresses of the piled up group initiate slip in the perfect lattice.

It follows then that we need consider only plastic flow due to dislocations from Frank-Read sources. If the material is brittle and plastic flow does not occur round the piled up group of dislocations, then the sources must be locked so as to be unable to operate. This locking may arise either by the segregation of impurities to the dislocations (Cottrell locking), or because some slip systems are inherently difficult to operate (e.g. non-basal slip in the hexagonal close packed lattice). We consider these in turn.

*(i) Cottrell locking of dislocations*

Cottrell (1948) and Cottrell and Bilby (1949) have shown how impurities (and especially interstitial impurities such as carbon or nitrogen in iron), can segregate to the dislocations and anchor them. The dislocations can only move by leaving their atmospheres of impurities behind, and so a stress is required to pull the dislocations free: this stress is the upper yield point. Once the dislocations are free they can move more easily; the stress then falls after the yield point has been reached and the deformation continues at a lower stress (the lower yield point).

Yield points have been observed in a number of metals (zinc, Wain and Cottrell 1950, cadmium, Gibbons and Cottrell 1947, titanium, Churchman 1955,  $\alpha$ -brass, Ardley and Cottrell 1953), but they are especially prominent in those having a body-centred cubic structure; in addition to iron, molybdenum (Trury and Kraus 1936, 1937), tungsten (Bechtold and Shewmon 1954) and tantalum (Bechtold 1955) all show pronounced upper and lower yield points. The occurrence of such yield points is an indication of the strength of the Cottrell locking in these metals.



There are two reasons why interstitial atoms should interact particularly strongly with dislocations in the body-centred cubic structure. First, this lattice has smaller interstices than do other lattices so that an interstitial atom has to expand the lattice to a greater extent giving a corresponding larger energy of interaction. Secondly an interstitial atom in this structure not only produces a pure volume change, but also distorts the lattice, changing it locally from cubic to tetragonal; thus the interstitials will interact with the dislocations through the shear stresses as well as through the hydrostatic component of the stress.

If, then, Cottrell locking occurs, it will anchor the Frank-Read sources against any stress less than the upper yield point, and so can provide the rigid matrix round the piled up group which is needed if fracture is to occur. The Cottrell locking will also help produce fracture in another way: the most difficult step in operating a source will be to free it from its locking impurities; this however needs to be done only for the first dislocation loop produced by the source. Hence once a source has been freed, it will be easier to continue operating this source than to free another; and so a small number of sources will operate, each producing many dislocations. Cottrell locking will thus lead to the formation of the large piled up groups needed for fracture. This conclusion is confirmed by the observation of course slip (fig. 4, Pl. 14).

The view that the brittle behaviour of these metals depends on the dislocations being locked by impurities is supported by the observation of Pratt (private communication) that if a single crystal of iron is deformed slightly at room temperature (so as to displace the dislocations and free them from their impurity atoms) and then cooled to liquid air temperature, the metal is still ductile though tested in an orientation in which brittle fracture would normally occur. Smith and Rutherford (1957) have found that brittleness is eliminated in removing impurities, and zone refined iron could be deformed at  $4.2^{\circ}\text{K}$  to an 80% reduction in area before fracture. Earlier work by Wain and Henderson (1953) had shown that the room temperature brittleness of chromium could be overcome by purification.

#### (ii) *Geometrical limitations on slip*

In metals with a close packed hexagonal structure, slip occurs primarily in the basal plane, and this provides only two independent slip systems (corresponding to two of the three slip directions in the basal plane). This is not sufficient to allow a polycrystalline aggregate to deform and maintain coherence between the grains, for von Mises (1928) has shown that to obtain an arbitrary plastic strain five independent slip systems are needed. Nor does twinning appear to be sufficient to provide the necessary additional degrees of freedom; for example, in magnesium the maximum strain obtainable from twinning is 7% for optimum orientation and will be much less for randomly orientated grains. At high temperatures, pyramidal and prismatic slip occur and can provide

the additional slip systems needed, but these are not operative at low temperatures. If then one grain yields, the surrounding grains may be orientated so as to be unable to deform in a way which will accommodate the slip in the first grain. Thus we can obtain one deforming grain in a rigid matrix and slip in this grain can proceed until the piled up groups of dislocations formed on its boundary are large enough to initiate a crack.

In these metals, we expect to see in the deforming grain not a single coarse slip line but a number of fine slip lines. However, a crack can just as well be initiated by the cooperative effect of several smaller piled up groups at the ends of a cluster of fine slip lines, provided that the total number of dislocations is still given by eqn. (9), (Stroh 1955 a). The relation between grain size and fracture strength is unchanged.

On this view then the brittle behaviour of hexagonal metals depends on the limited number of slip systems available. The way in which the operation of additional slip systems can cause the material to be ductile is very clearly shown by Hauser *et al.* (1957) by observations on  $\alpha$  solid solutions of lithium in magnesium. The addition of lithium to magnesium causes prismatic slip (10 $\bar{1}$ 0), [2 $\bar{1}$ 10] to occur at progressively lower temperatures and for the alloys richest in lithium (10.4 and 14.5 atm %) prismatic slip was observed at temperatures as low as 4.2°K. At the same time the ductility increased markedly, and the 14.5% alloy necked and broke in a typically ductile manner at all temperatures investigated.

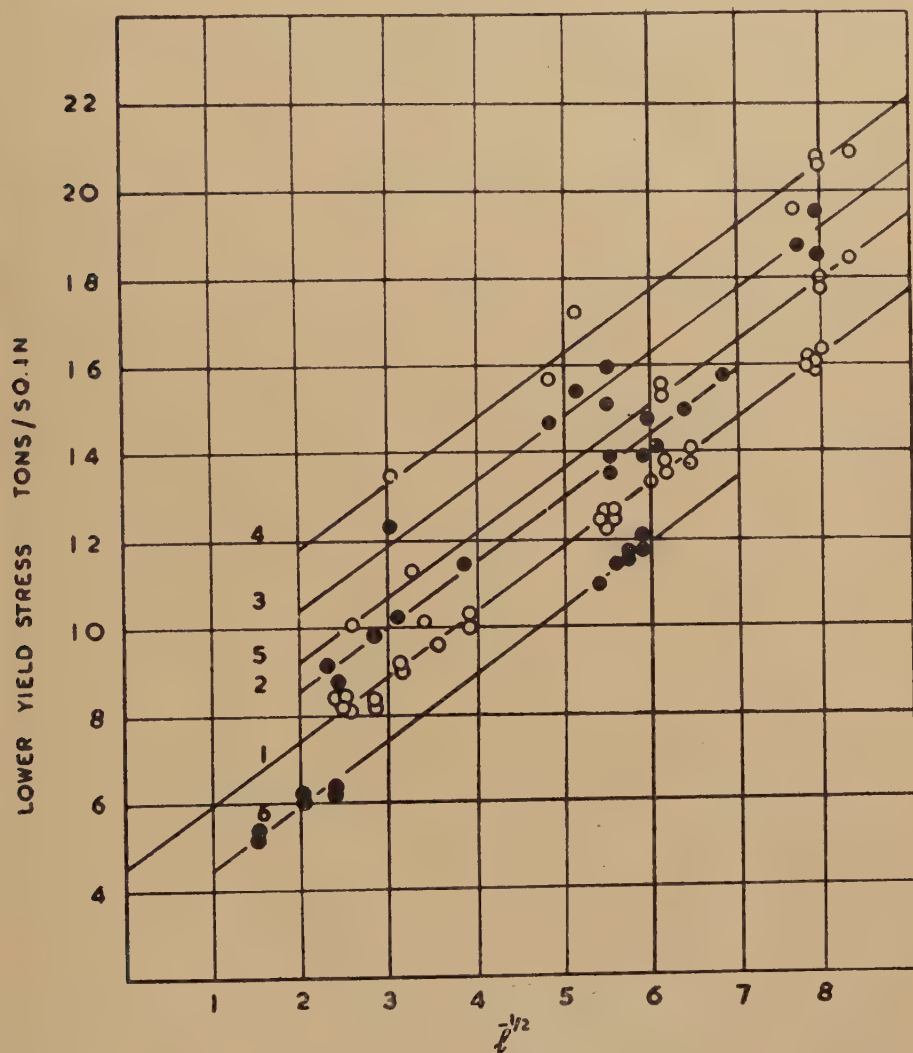
In considering why prismatic slip should occur, it is to be noted that the addition of lithium decreased the axial ratio from 1.624 (pure Mg) to 1.610 (Mg, 14.5% Li); this latter value is nearer the axial ratios of titanium (1.59), zirconium (1.59), and beryllium (1.57) in all of which prismatic slip is normal. Hauser *et al.* suggest that the Peierls-Nabarro force needed to move the dislocations depends on both the slip plane and the axial ratios, and, for sufficiently small axial ratios, becomes less on the prismatic planes than on the basal plane. If the Peierls-Nabarro force is large the dislocations will be unable to move except with the help of thermal fluctuations so that this slip system will be observed only at high temperatures.

#### § 4. THE LOWER YIELD POINT

In a metal such as iron where Cottrell locking occurs, thermal vibrations will help free the dislocations from their locking impurities and so they will be less firmly bound at high temperatures than at low. At a sufficiently high temperature the stresses of the piled up group will be able to free the dislocations near it, and then slip will continue into the next grain instead of fracture occurring and so the metal will be brittle at low temperatures, ductile at high. The stress at which the slip propagates through the material, if it is ductile, is the lower yield point: we shall consider the nature of the lower yield point in the present section and postpone till later (§ 5) a discussion of the conditions determining the transition from brittle to ductile behaviour.

As Petch (1953) has emphasized the lower yield and the cleavage strength depend in the same way on the grain size, suggesting that similar mechanisms operate in the two cases. Figure 12, due to Cracknell and Petch (1955), showing the relation between the lower yield point and the grain size of mild steel may be compared with fig. 6; similar results on the lower yield point have been obtained by Hall (1951) and Petch (1953). Both cleavage and yielding may be interpreted as due

Fig. 12



The relationship between lower yield point and grains size for (1) annealed mild steel *En 2*; (2) *En 2* nitrided; (3) *En 2* quenched from 650°C; (4) *En 2* quenched, aged one hour at 150°C; (5) *En 2* quenched, aged 100 hours at 200°C; (6) annealed Swedish iron. Cracknell and Petch (1955).



to the stresses round a piled up group ; cleavage, to the tensile stress, and yielding to the shear stress.

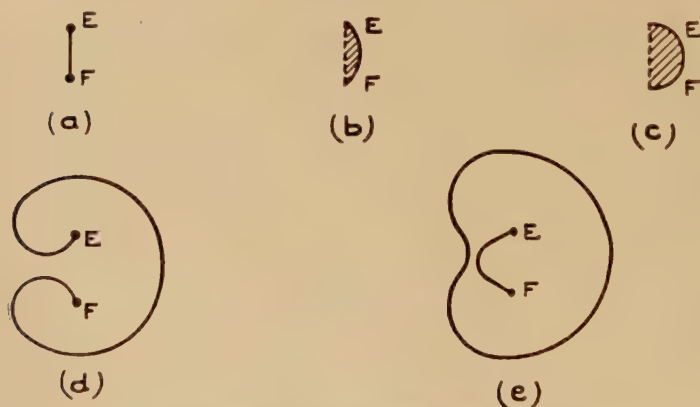
At a distance  $r$  from the head of a piled up group the shear stress is

$$\sigma = \beta \sigma_s (L/r)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where  $\beta$  is a numerical factor, approximately unity but depending on the orientation. (Equation (13) for the shear stresses corresponds to eqn. (5) for the normal stresses.)  $r$  must be taken equal to the distance from the piled up group of the source operated. Now fig. 12 shows that a number of quite different specimens of steel all give lines with the same slope ; hence this slope cannot depend on any quantity such as the mean spacing of the dislocation network which would vary considerably from one specimen to another. We therefore suppose that there will occur a number of sources of various lengths and at various distances from the piled up group ; of these the most favourably situated ones will automatically be selected and operated first, and these will always be the same distance (i.e. for different specimens but at a given temperature) from the piled up group. The reason why some sources are favoured above others is as follows.

For a source to operate, the stress on it must, firstly, be great enough to pull it free from its locking impurities ; for this the stress must be at least equal to the upper yield point  $\sigma_u$ . Secondly, when the source has been freed the stress must be sufficient to bow it out in the characteristic manner of the Frank-Read source shown in fig. 13 ; for this a stress

Fig. 13



The operation of a Frank-Read source.

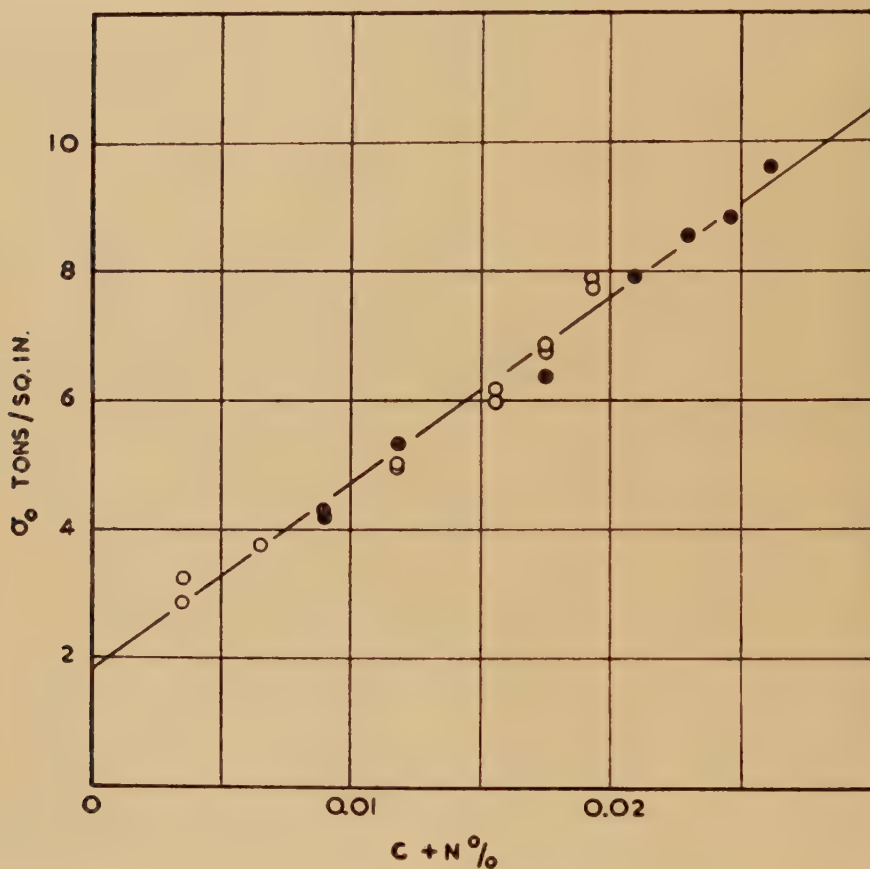
of  $Gb/l$ , where  $l$  is the length of the source, is needed. It is clear from fig. 13 that a source of length  $l$  needs an area whose linear dimensions are at least  $l$  to operate in, and so the source cannot be nearer than a distance  $l$  to the piled up group ; in general the nearest sources of length  $l$  will be a distance  $\alpha l$  from the piled up group, where  $\alpha$  is a constant of order, but greater than, unity. Then the sources very near the piled up



#### 4.1. *The Stress Needed to Move a Dislocation*

We have seen that, in order to obtain a relation of the correct form between the cleavage strength, or the lower yield point, and the grain size, it is necessary to introduce the idea of a frictional force opposing the motion of the dislocations. Figure 12 shows clearly that this frictional force,  $\sigma_0$ , which is measured by the intercept on the stress axis, differs considerably for different specimens; this is in direct contrast to the behaviour of the gradients  $K$  and  $K^*$ , which are the same for all

Fig. 14



Effect of the concentration of  $C+N$  on  $\sigma_0$ . Cracknell and Petch (1955).

specimens. Cracknell and Petch (1955) suggest that  $\sigma_0$  is due to internal stresses around fine precipitates or randomly distributed solute atoms; the variation of  $\sigma_0$  shown in fig. 12, on nitriding or quenching (which alters the concentration), and on quench ageing (which alters the state of dispersion), is in accord with this idea. They base a quantitative treatment of the effect on the work of Mott and Nabarro (1948) and



Mott (1952) which leads to the result that  $\sigma_0$  should vary linearly with the concentration of carbon and nitrogen over the range of concentrations used, a prediction which fig. 14 shows is confirmed by observation ; the predicted slope, however, is rather less than that observed, probably owing to approximations which have had to be made in the theory.

The value of  $\sigma_0$  which remains at zero concentration has been interpreted by Cracknell and Petch as the Peierls force on the dislocations. If this is correct it means that the Peierls force in  $\alpha$ -iron (and probably in other body-centred cubic metals) is rather larger than has hitherto generally been supposed. Heslop and Petch (1956) have found that the part of  $\sigma_0$  which is independent of the concentration is strongly temperature dependent ; this is consistent with its interpretation as a Peierls force.

## § 5. THE BRITTLE-DUCTILE TRANSITION

### 5.1. *The Probability of Brittle Fracture*

In the preceding sections the brittle fracture at low temperatures, and the general yielding at higher temperatures have been considered ; it remains to discuss the conditions which determine the transition from brittleness to ductility†.

By far the greatest number of experimental investigations have been done on iron and steels, and the experimental results which are quoted in this section all refer to these metals. In such cases, the locking of the dislocations is due to impurities, and the transition will be discussed in terms of this. However, the ideas used can be applied more generally, and any way of locking the dislocations from which they can be released under a stress by supplying an activation energy, would serve equally well. For example, the discussion of § 3.6 suggests that in hexagonal metals the brittleness is due to the locking of the dislocations in non-basal planes by a large Peierls force ; but the transition may still be treated as in the present section.

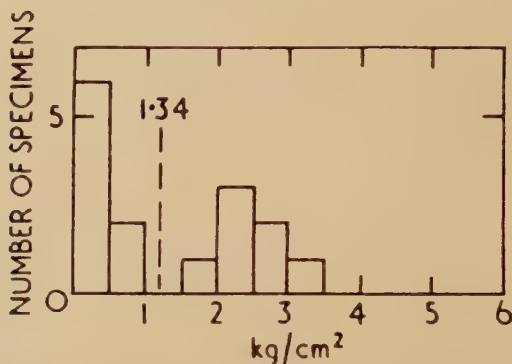
The transition is usually studied experimentally by impact test, such as the Charpy and Izod tests, in which the energy absorbed in breaking a specimen is measured. At low temperatures when the specimen is brittle this energy is small, but it is greater at temperatures for which the specimen is ductile. In the transition region a wide range of energy values are obtained ; a number of workers (Vanderbeck *et al.* 1953, Ulmo and Bastenaire 1953, Crussard *et al.* 1956) have shown that these values are distributed bimodally, as is apparent from the histogram shown in fig. 15 ; low energies characteristic of the brittle condition and high energies characteristic of the ductile both occur, but intermediate values

---

† A very full experimental investigation of this topic has recently been made by Crussard *et al.* (1956). They distinguish between the initiation transition considered here and the propagation transition corresponding to a change in the mode in which the crack spreads.

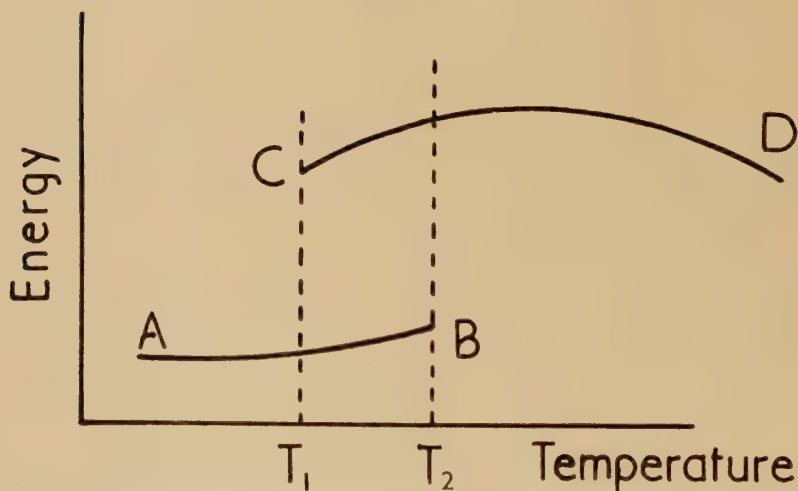
are exceptional. Thus the results of such a test may be represented ideally as in fig. 16. AB is the brittle curve and CD the ductile; below the temperature  $T_1$  all points lie on AB so that this is the brittle region, while above  $T_2$  there is the ductile region. The transitional region extends from  $T_1$  to  $T_2$ , and here points may lie on either curve according

Fig. 15



Histogram of mild steel specimens broken by slow bending at  $-140^{\circ}\text{C}$  (15 specimens). Crussard *et al.* (1956).

Fig. 16



Variation of fracture energy with temperature typically obtained in an impact test.

to a probability law depending on the temperature, so that the two types of fracture, ductile and brittle, may coexist at the same temperature. We wish to find the probability that the metal will be brittle at any given temperature.

Now according to the ideas we have developed, the dislocations are locked by Cottrell impurities. At high temperatures, thermal fluctuations will be able to free the dislocations round a piled up group and so relieve the stresses there; in this case, the metal will be ductile. On the other hand, at low temperature, the thermal fluctuations will be insufficient to free the dislocations before the piled up group has grown large enough to generate a crack; then the metal is brittle. We therefore identify the probability of brittle fracture with the probability that the dislocations near a piled up group will not be released.

To free a dislocation, an activation energy  $U(\sigma)$ , depending on the stress  $\sigma$ , must be supplied, and so the probability per unit time of the dislocation being freed may be written  $\nu \exp \{-U(\sigma)/kT\}$ , where  $\nu$  is a constant of the dimensions of a frequency. Hence the mean time† elapsing before a dislocation is released is

$$\tau = (1/\nu) \exp \{U(\sigma)/kT\}. \quad . \quad . \quad . \quad . \quad . \quad (17)$$

The release of a dislocation will be a purely random event, subject only to the restriction that it occurs on the average after time  $\tau$ . Such random processes are familiar in many branches of physics, e.g. radioactive decay, and the probability that the event, here the release of the dislocation, will *not* have occurred at time  $t$  is

$$p = \exp (-t/\tau),$$

or substituting  $\tau$  from eqn. (7),

$$p = \exp [-\nu t \exp \{-U(\sigma)/kT\}]. \quad . \quad . \quad . \quad . \quad (18)$$

As the piled up group is formed the stress  $\sigma$  which it produces on any dislocation will increase from zero to some maximum value. The probability of the dislocations being released while the stress is still small is negligible, and so in eqn. (18)  $\sigma$  must be taken as the maximum value of the stress;  $t$  will be a measure of the time during which the stress is near this value. Then eqn. (18) will give the probability of brittle fracture at any temperature. Because of the double exponential,  $p$  will change fairly abruptly from 1 to 0 as  $T$  increases, so that the model reproduces the sharp transition observed. The transitions will occur at a temperature

$$T_c = U(\sigma)/k \log \nu t. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The activation energy  $U(\sigma)$  needed to free a dislocation depends on the stress  $\sigma$  acting on it, and this in turn depends on the distance of the dislocation from the piled up group. Thus in general there will be a wide range of activation energies. However the consideration of

---

† Equation (17) gives the relation between the time to release the dislocations and the stress. The existence of such a relation is just the interpretation in dislocation theory of the experimental result (e.g. Taylor 1946) that the observed yield stress depends on the duration of loading.



§ 4 led to the conclusion that the Frank-Read sources which were most easily operated would occur at some definite distance from the piled up group, and it is the operation of these sources which will determine whether the metal is ductile or brittle. Accordingly the value of the activation energy which is of importance is that needed to free one of these most easily operated sources when the piled up group has grown just large enough to initiate a crack. This activation energy has a uniquely determined value, and it is to this value that we shall refer in speaking of the activation energy.

Stroh (1955 b) has attempted to estimate  $T_c$  by using Cottrell and Bilby's calculations (1949) of  $U(\sigma)$ ; though a transition temperature of the correct order was obtained, there are so many sources of uncertainty involved that little weight can be attached to the final result.

It seems better to compare equation directly with experiment. If a number of specimens are broken at the same temperature, they may be classified into two groups according as they fail in a brittle or a ductile manner; the energy absorbed in fracture provides a useful criterion for this (fig. 15). If this is done  $p$  may be identified with the fraction of the total number of specimens which break brittly at each temperature. Equation (18) may be rewritten

$$\log \log (1/p) = \log vt - U/kT, \quad . \quad . \quad . \quad (20)$$

or  $\log \log (1/p)$  is a linear function of  $1/T$ ; the results of Vanderbeck *et al.* (1953)† shown in fig. 17 are seen to follow this law within the experimental scatter. From the slope of the line, we obtain the activation energy  $U = 0.28$  ev. This value has been entered in table 3, together with estimates of the activation energy obtained by treating the results of Baeyertz *et al.* (1949) and of Ulmo and Bastenaire (1953) in the same way. These values will be discussed later (§ 5.4) after we have considered other properties of the transition also leading to an estimate of the activation energy. Meanwhile we may note that in order of magnitude the values obtained are not inconsistent with those expected from the theoretical calculations of Cottrell and Bilby (1949).

### 5.2. Effect of Strain Rate

In eqn. (18) describing the brittle-ductile transition, the only quantity which will depend on strain rate is  $t$ , the time for which the stress on a dislocation is near the value  $\sigma$ . We expect  $t$  to be given apart from a numerical factor, by

$$t \simeq \sigma \dot{\sigma}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where a dot over a letter denotes differentiation with respect to time. The stress  $\sigma$  which the piled up group produces will increase in proportion

---

† Vanderbeck *et al.* have fitted a Galton curve, obtained by integrating a Gaussian distribution, to their experimental points; however, this is not based on any definite atomic mechanism.



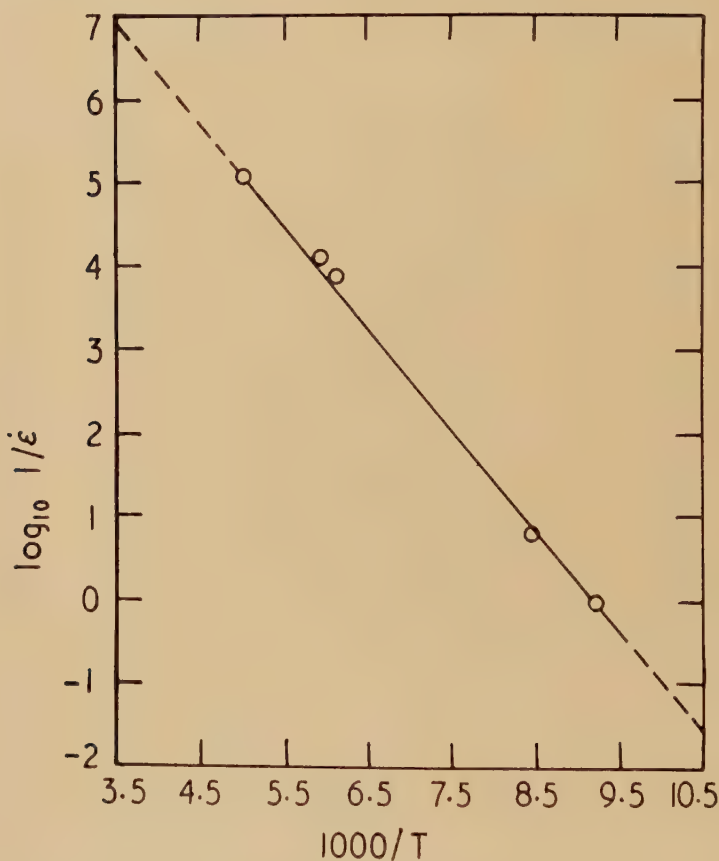
Inserting this value of  $t$  into eqn. (19), we find that the relation between transition temperature and strain rate is given by

$$1/T_c = -(k/U) \log \dot{\epsilon} + C, \quad . . . . . (23)$$

where  $C$  is independent of  $T_c$  and  $\dot{\epsilon}$ .

Such a linear relation between  $1/T_c$  and  $\log \dot{\epsilon}$  has been observed by Wittman and Stepanov (1939) whose results are reproduced in fig. 18,

Fig. 18



Dependence of transition temperature on strain rate for steel. Wittman and Stepanov (1939).

and similar results have been obtained by MacGregor and Grossman (1952). The slopes of these lines also lead to an estimate of the activation energy  $U$ , and the values obtained are shown in table 3,



Table 3. Values of the Activation Energy needed to Free a Dislocation in Steel.

Type of Observation	Activation energy U ev	Authority
Probability of brittle fracture	0.28 0.19 0.36	Vanderbeck <i>et al.</i> (1953) Ulmo and Bastenaire (1953) Baeyertz <i>et al.</i> (1949)
Strain rate	0.24 0.44	Wittman and Stepanov (1939) MacGregor and Grossman (1952)
Variation of grain size	(a) 0.30 (b) 0.22 }	Hodge <i>et al.</i> (1949)

### 5.3. Effect of Grain Size

It is generally agreed that a fine grained specimen has a lower transition temperature than a coarse grained one. On our model, there are several reasons why that should be so. Firstly it follows immediately from the considerations of § 5.2 that  $t$ , the effective time for which the Frank-Read sources round the piled up group are stressed, depends on the grain size  $d$ ; eqn. (22) shows that  $t$  is proportional to  $d^{-1/2}$ . This in turn leads to a dependance of the transition temperature on  $d$ ; though only about 15% of the observed variation can be accounted for in this way.

A more important contribution arises because the grains are not all identical but differ in size and orientation, so that some yield more easily than others. Slip will, on the model, propagate by a piled up group in one grain operating sources, and so starting slip, in the next; a grain is ductile if it yields before the piled up group in the previous grain to slip has grown large enough to initiate a crack. Now suppose that the temperature and strain rate are such that the *average* grain is just ductile. At any stage in the deformation there will be a number of grains which may slip next; the one to do so will be that one which yields most easily. As long as there is at least one grain ready to slip, the deformation can continue and the material as a whole will be ductile. Occasionally, however, it will chance to happen that all the grains through which slip can propagate yield only with difficulty; then the applied stress must increase slightly and the piled up groups will grow larger; this may be sufficient to initiate a crack so that the material, as a whole, is brittle. Now the probability of finding a grain yielding at any time is the product of the number of grains considered and the probability  $\nu \exp(-U/kT)$  of a given grain yielding. If specimens of the same geometrical size and shape are considered the number of grains in any

region is inversely proportional to the volume of a grain and so proportional to  $d^{-3}$ . Thus the frequency  $\nu$  must be multiplied by a factor proportional to  $d^{-3}$ ; it is convenient to redefine  $\nu$  so that it includes this factor; the transition temperature is then still given by an expression of the form (19).

The two effects we have been considering give the product  $\nu t$  proportional to  $d^{-7/2}$ ; eqn. (19) then gives the relation between transition temperature and grain size as

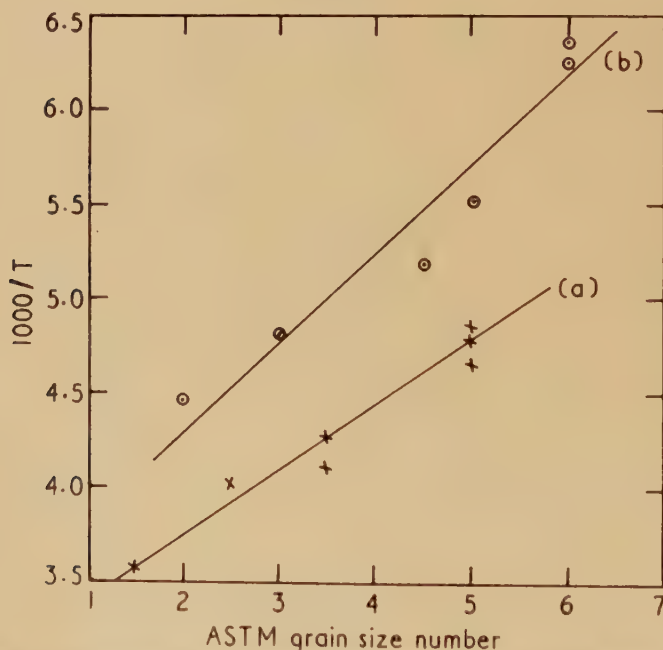
$$1/T_c = -\frac{7}{2} (k/U) \log d + C, \quad . . . . . (24)$$

where  $C$  is independent of  $T_c$  and  $d$ . Since the conventional grain-size numbers (ASTM or Jernkontoret) form a logarithmic scale, an increase of unity in the grain-size number  $N$  doubling the number of grains per unit area, eqn. (24) may be written

$$1/T_c = 1.21(k/U)N + C'. \quad . . . . . (25)$$

Hodge *et al.* (1949) have concluded that there is a linear relation between transition temperature and grain-size number; however their results, which are replotted in fig. 19 agree equally well with eqn. (25).

Fig. 19



Dependence of transition temperature on grain size. Steel with (a) 0.03% Ni and (b) 3.64% Ni. Experimental points of Hodge *et al.* (1949).

The slope of the lines give other estimates of the activation energy  $U$ , which have again been entered in table 3.

#### 5.4. The Activation Energy

In the preceding sections we have seen how the activation energy  $U$  may be found from a number of quite different types of observation. The estimates of  $U$  which have been obtained are collected in table 3; there is a marked tendency for values near 0.3 eV to occur.

Part of the variation in values obtained is doubtless due to experimental uncertainties which in some of the cases may be important. It is unlikely, however, that this accounts for the whole of the variation, and indeed there seems no justification for expecting all the entries in the table to be identical; they refer to experiments made on different steels having different compositions. The presence of small amounts of alloying elements is known to produce changes in the transition temperature (cf. for example Rinebolt and Harris 1951); though the mechanism by which this is brought about is not as yet understood, the effect may well be due, in part, to a change in the activation energy needed to free the Frank-Read sources, which is caused by such impurity atoms. For this reason it would be of great interest if the different types of observation we have been considering were to be repeated on specimens of the same composition so that the activation energies deduced would be strictly comparable.

Bearing in mind the variation in composition, we see that the various activation energies obtained agree as well as may be expected; in particular there is no evidence for a systematic difference in the activation energies obtained from observations of different type. This general agreement, and also the fact that the value obtained (about 0.3 eV) is consistent with Cottrell and Bilby's theoretical curves (1949), is evidence in favour of the model of fracture which has been developed.

#### 5.5. Notch Brittleness

The presence of a notch in the specimen has an important effect on fracture and produces a large increase in the transition temperature. In considering how this is brought about, we note the following properties of the notch.

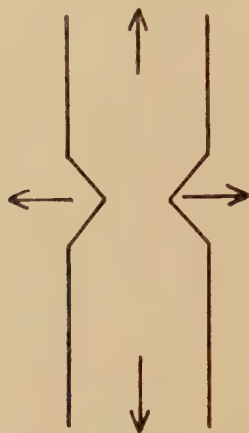
(i) If yielding has not yet occurred and the material is elastic, the notch will act as a stress concentrator so that large stresses occur near the notch. The smaller the radius of curvature of the root of the notch, i.e. the sharper the notch, the greater is the stress magnification produced, but these large stresses occur only in a relatively small region near the notch root.

(ii) Consider a notched specimen, subjected to a uniaxial tension (fig. 20). Yielding in the notch tends to produce a contraction in the cross section there, which is opposed by the unyielding material on either side; thus a transverse tension will be produced. The effect of the notch then is to change the stress system from a uniaxial to a biaxial, or triaxial, one. Since the maximum shear stress is equal to half the



difference of the principal stresses, the shear stress will be reduced by the presence of the notch, while the maximum normal stress is unchanged. Plastic flow depends on the shear stresses, and so the axial stress needed to produce yield is increased by the presence of the notch; according to Prandtl (1923), Orowan (1945) and Hill (1949), the axial yield stress is increased by a factor of about 3.

Fig. 20



Stress system near a notch.

(iii) Near the notch the strain rate will be enhanced. For elastic strains this follows from (i) on account of the proportionality between stress and strain. When yielding occurs the flow will be such as to relieve the stresses present; thus the elastic stress concentration near the root of the notch is converted into a region of plastic strain concentration, and hence there will also be a high plastic strain rate near the notch.

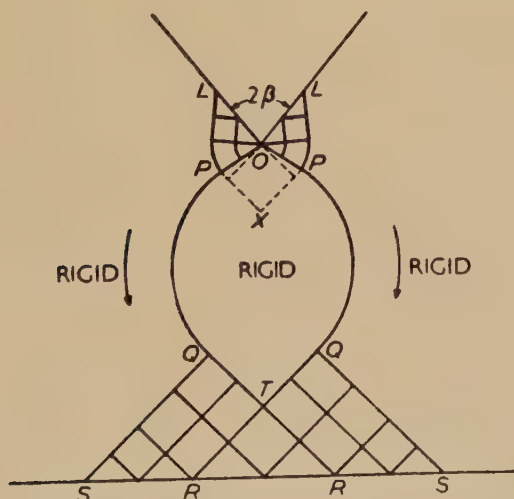
The classical theory of notch brittleness which originates from the work of Mesnager (1906) and Ludwik and Scheu (1923) and which has subsequently been elaborated by Orowan *et al.* (1945) and Orowan (1945), takes (ii) as the principal effect in providing brittleness. Fracture is assumed to depend on the tensile stress, so the triaxial stress system produced by the notch will tend to favour fracture rather than yielding.

On the present model this will not be the case, for both fracture and yielding depend on the shear stress and so will be affected in the same way by a change in the nature of the stress system. There is some experimental evidence that this is the case. McAdam *et al.* (1947) slightly deformed notched specimens at room temperature (where they were ductile) so that some plastic flow took place round the root of the notch and the stress concentration considered in (i) above was eliminated; the triaxial stress system (ii) of course, still remained. On reloading the specimens at a lower temperature, it was found that the fracture

stress was *increased* by the notch in the same manner as the yield stress. From their experiments on the propagation of brittle cracks, to be discussed more fully in § 6.3, Felbeck and Orowan (1955) conclude that the triaxial nature of the stress is unimportant and that the decisive factor in determining whether brittle or ductile fracture will occur is the velocity dependance of the yield stress. Other evidence supporting the view that fracture depends on the shear stress rather than the tensile stress has been reviewed by Zener (1948).

The view taken here is that the increase in strain rate mentioned in (iii) is the principal effect of the notch in causing brittleness. It was seen in § 5.2 that an increase in strain rate would increase the transition temperature; while Castro and Gueussier (1949) and Crussard *et al.* (1956) have emphasized that the effects of temperature, strain rate, and triaxiality of the stress are equivalent. The importance of the enhanced strain rate near a notch has been emphasized by Holloman (1944) who estimates that the elastic strain rate in a notch bend test may be as much as  $10^7$  times the strain rate in a normal tensile test; this alone is sufficient to produce a change in the transition temperature of the order of  $100^\circ\text{C}$ . When plastic flow occurs at the notch the strain rate will be even greater so that a further increase in the transition temperature will occur. Though a number of studies of the initial yielding in notched specimens have been made (Green and Hundy 1956, Crussard *et al.* 1956) and good agreement found between theoretical and experimental slip line fields (figs. 21 and 22, Pl. 15) it has not as yet proved possible to make any quantitative estimate of the maximum plastic strain rate

Fig. 21



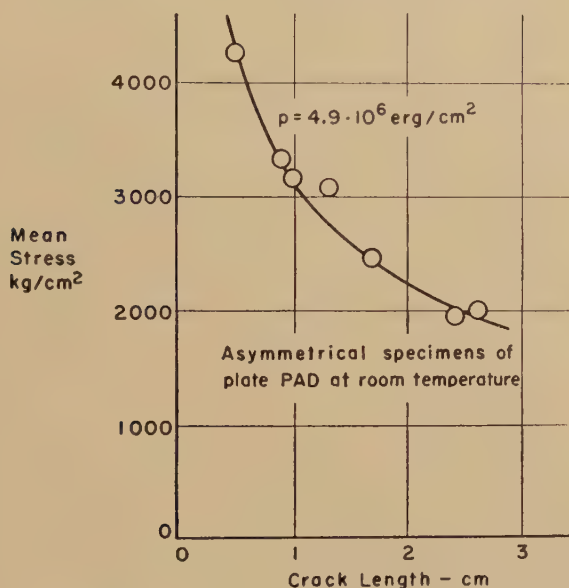
Theoretical slip line field for pure bending with a V notch. The central pivot OPQTQPO remains rigid and the parts of the bar on either side rotate about this (compare fig. 22, Pl. 15). Green and Hundy (1956).





In § 2.2 it was seen that the Griffith condition for the growth of a crack could also be obtained by requiring that the stress at about an atomic distance from the tip of the crack should equal the molecular cohesion of the material. Orowan (1955) has also considered what must be the corresponding interpretation in the present case. We must now consider macroscopic quantities only, since clearly the treatment will no longer apply at distances less than the thickness  $t$  of the plastically deformed layer, i.e. 0.2 to 0.4 mm. At a distance  $t$  from the tip of the crack the stress will be magnified by a factor of about  $(c/t)^{1/2}$ ; for the crack of fig. 24, this gives a stress of about 6000 kg/cm<sup>2</sup>, or about 40 tons/in.<sup>2</sup>,

Fig. 24



Variation of stress needed to propagate a crack, with crack length.  
Felbeck and Orowan (1955).

which is of the same order as the observed cleavage strength of steel. An indication as to the mechanism operating is afforded by Tipper's observation (1945, 1957) that the crack does not propagate continuously; rather a number of small unconnected cracks are formed in grains ahead of the main crack, and these subsequently join up (fig. 25, Pl. 16). For the crack to propagate by such a process of repeated re-initiation of fracture, it is understandable that the macroscopic stress should have to attain locally the observed cleavage strength.

The cracks in the individual grains might be produced by piled up groups of dislocations at the ends of slip lines as has been considered in § 3; alternatively a crack in a previously fractured grain could serve instead of the piled up group, for, as was remarked before, the stresses

round a crack and a slip line are very similar. In view of the high speed at which a brittle crack propagates the latter seems more likely as requiring less time to operate ; but in either case the same macroscopic stress is needed.

Our picture of the propagation of a crack is then as follows. The main crack produces locally, but on a macroscopic scale, a region of stress concentration at its tip. In this region there are microcracks, due to cleavage within grains, ahead of the main crack. The stress acting on these microcracks, will be magnified again to produce, now on an atomic scale, a stress equal to the atomic cohesion of the material ; in this way, the stress is raised sufficiently to initiate cleavage in other grains. Finally the microcracks will link together and join up with the main crack by tearing, and it is during this process that most of the plastic work deformation associated with the fracture is done.

In addition to producing plastic deformation, the propagation of the crack may also induce twinning as, for example, the Neumann lamellae observed in iron (Tipper and Hall 1953). The stresses associated with a moving crack have been studied by Bilby and Bullough (1954) with special reference to the formation of twins ; they found that they could give a consistent explanation of the formation of the twins and kinkbands which Deruyttère and Greenough (1954) had observed in the cleavage of zinc single crystals. All such processes will result in the absorption of some energy, and this must be added to the plastic work  $p$  introduced above.

### 6.2. *The Velocity of a Crack*

Mott (1948) has considered the speed at which a Griffith crack will spread. He points out that when the crack is moving the kinetic energy of the material disturbed by the crack must be taken into account ; for a crack of length  $c$  moving with speed  $V$  through a material of density  $\rho$ , the kinetic energy may be written as  $k\rho c^2 V^2 \sigma^2 / E^2$ , where  $k$  is a numerical constant. Adding this term to the Griffith eqn. (2), the condition  $dW/dc=0$  for energy balance gives the velocity of the crack to be

$$V = (\pi/4k)^{1/2} (E/\rho)^{1/2} (1 - c_G/c), \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where  $c_G$  is the Griffith length

$$c_G = 4\gamma E / \pi \sigma^2. \quad . \quad . \quad . \quad . \quad . \quad (28)$$

When the length of the crack is large compared with the length  $c_G$ , eqn. (27) shows that the velocity will tend towards the limiting value  $(\pi/4k)^{1/2} (E/\rho)^{1/2}$  ;  $(E/\rho)^{1/2}$  is here the velocity of the longitudinal waves in a thin rod of the material. Roberts and Wells (1954) have evaluated the constant  $k$  and so obtained the value of  $0.38 (E/\rho)^{1/2}$  for the limiting velocity.

The treatment just considered strictly holds only when the velocity is small as 'relativistic' effects which occur when the velocity of the crack approaches that of sound have been ignored. An exact treatment

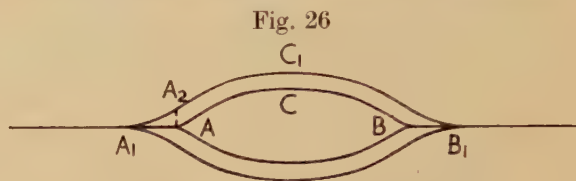
Table 4. Measured Velocities of Cracks

Investigator	Material	Velocity $V_0 = \sqrt{(E/\rho)}$ of longitudinal waves ft./sec.	Velocity $V$ of crack ft./sec.	$V/V_0$
Schardin and Struth (1938)	Glass	17000	4650-5150	0.27-0.30
	Fused Quartz	17000	7200	0.42
Hudson and Greenfield (1947)	Steel	16500	3370	0.20
Barstow and Edgerton (1939, 1941)	Glass	18000	5000	0.28
Smith and Ferguson (1950)	Cellulose acetate	3700	1000-1375	0.27-0.37
Kennedy (1945)	Steel	16500	4500	0.27
Boodberg <i>et al.</i> (1948)	Steel	16500	4600-6600	0.28-0.40
Robertson (1953)	Steel	16500	6000	0.36
Bueche and White (1956)	Silicone Rubber I	67	11-26	0.16-0.38
	Silicone Rubber II	150	26-53	0.18-0.35
	Irradiated polyethylene	380	150-300	0.38-0.78
	Plexiglas	1600	330	0.21
	Glass	19000	5000	0.26
	Fused silica	24000	7300	0.32



in this case presents a rather intractable problem; but assuming the result that the velocity of the crack tends to a limiting value independent of the Griffith length, and hence of the surface energy, this limiting velocity is easily seen to be just the velocity of the Rayleigh surface waves.

For suppose the crack at any instant of time extends from A to B (fig. 26) and consider the stress acting across  $AA_1$  just ahead of the crack.



To illustrate the spread of a crack.

When the crack moves from A to  $A_1$  an amount of work is done which is given by the product of the stress on  $AA_1$ , and the displacement  $AA_2$ ; this work must be equal to the surface energy of the new surface  $A_1A_2$ . But the limiting velocity of the crack is independent of the surface energy, and so this surface energy may be taken to be zero. Hence, since  $AA_2$  is not zero, the stress on  $AA_1$  must be zero. Thus the propagation of the crack involves the surface  $A_1ACBB_1$  deforming to  $A_1A_2C_1B_1$ , the surface being unstressed throughout, and so is just a disturbance moving on a stress-free surface. The velocity with which such a disturbance moves is the velocity of the Rayleigh waves.

If Poisson's ratio is taken to be  $\frac{1}{4}$ , the velocity of the Rayleigh waves (Love 1927) is  $0.919\sqrt{(G/\rho)} = 0.58\sqrt{(E/\rho)}$ . This value is somewhat higher than that obtained by Wells and Roberts quoted above; the difference is probably due to the neglect of relativistic effects by these authors.

A number of measured values are collected in table 4. The values are seen to be all a little less than the theoretical value: this is not altogether unexpected as any dissipative process must reduce the crack velocity. Moreover, Tipper (1957) suggests that cracks propagate in steel in a jerky manner, coming to rest and then moving forward again; such a process would give a mean crack velocity rather less than the limiting velocity.

### 6.3. The Propagation Transition

A crack can propagate in a brittle manner only if the dislocation sources near which it passes are locked so that plastic flow is inhibited. The locking of the sources may be due to impurity atoms. At higher temperatures thermal fluctuations will be able to free the dislocations so that the material is ductile, and we obtain a transition temperature for the propagation of the crack in the same way as we obtained one for the initiation. The only change needed in the treatment of § 5.1 is

that  $t$  is now the time the crack takes to pass near a Frank-Read source : this will be of order  $r/V$ , where  $r$  is the distance of closest approach of the crack. Such transition temperatures have been measured by Robertson (1951, 1953). Specimens of ship plate were subjected to a uniform tensile stress, and a temperature gradient in a direction perpendicular to the stress axis was maintained by heating at one end and cooling with liquid nitrogen at the other. A cleavage crack was started from a saw cut in the cold end by a small explosion. The crack travelled some distance in the direction of increasing temperature, and the temperature at the point where it came to rest was noted.

If any plastic flow occurs, this will involve the expenditure of energy, and so will tend to reduce the speed of the crack. On the other hand if the crack slows down, the dislocations will be subjected to the stress field of the crack for a longer time and so more plastic flow will occur†. The process is thus a cumulative one, and if plastic flow starts it may build up rapidly until the crack is no longer able to propagate : in this case the material is ductile.

The way in which the velocity of the crack determines whether it shall propagate in a brittle or a ductile manner is very clearly shown in Felbeck and Orowan's experiments (1955) in which a crack which had been introduced artificially in a low carbon steel was caused to propagate. In fig. 27, Pl. 16, 1 is the initial position of the crack. When a load was applied at first considerable plastic deformation occurred at the tip of the crack which started to propagate as a ductile crack 2. After a short run, this changed into a brittle crack 3 which spread at high velocity. Owing to an inability of the testing machine to maintain the load during the rapid extension of the specimen, the load dropped drastically (to about half its initial value) and the crack stopped in the material. On increasing the load again the process could be repeated, the crack starting as a ductile crack 2 and then becoming brittle 3. The contraction of the thickness of the plate due to plastic deformation when the crack is ductile can be seen in the photograph. The only simple explanation of the behaviour of the crack is that brittle propagation cannot occur until the velocity of the crack has attained some minimum value.

A simple model which exhibits some of the features observed has been proposed by Gilman (1956). Equation (27) for the velocity  $V$  of a crack may be rewritten

$$V = V_0(1 - U_s/U_E)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

where  $V_0$  is the limiting velocity of a truly brittle crack,  $U_s$  is the surface energy and  $U_E$  the elastic energy of the crack. If there is any plastic deformation at the fracture surfaces, then  $U_s$  must, as in § 6.1 include the plastic work done. We now assume that the amount of plastic work

---

† The relation between the amount of plastic work done and the velocity of the crack has been considered by Hall (1953).

is proportional to the time for which the stress is applied, and so inversely to the velocity  $V$  of the crack ; that is, we write

$$U_s = U'_s / V,$$

eqn. (28) then becomes

$$V = V_0(1 - U'_s/U_E V)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad (29)$$

If  $U'_s > (2/3\sqrt{3})U_E V_0$ , eqn. (29) is not satisfied by any real value of  $V$  ; that is, if plastic flow occurs too copiously, the propagation of a crack in a quasi-brittle manner is impossible. By assuming  $U'_s$  to be an increasing function of the temperature, we thus obtain a critical temperature above which the brittle propagation of a crack does not occur. For  $U'_s < (2/3\sqrt{3})U_E V_0$ , eqn. (29) is satisfied by a value of  $V$  between  $V_0$  and  $V_0/\sqrt{3}$ , depending on the value of  $U'_s/U_E$  ; this represents the high velocity brittle crack which is obtained when plastic flow is restricted. Thus far the model reproduces fairly well the observed features of crack propagation. However under the conditions in which brittle propagation is possible, eqn. (29) is also satisfied by a second and smaller value of  $V$  ; for small values of  $U'_s$ , this value of  $V$  is approximately  $U'_s/U_E$ , and becomes very small if plastic flow is much restricted. Such a low velocity brittle crack is *not* observed.

It should be noted that the assumption on which the model is based, that the amount of plastic work done is proportional to the duration of loading, is certainly not true in general. This is shown by the experiments of Clark and Wood (1949) on delayed yielding in iron. If a fixed stress is applied to the material, at first little plastic flow occurs, but after a certain time catastrophic yield sets in ; this delay time is just the time needed for thermal fluctuations to release the dislocations from their locking impurities.

#### 6.4. *The Effect of Transverse Stress*

Robertson (1953) has found that the temperature at which his cracks were arrested was independent of the stress transverse to the crack provided this stress was above some minimum value. For stresses below this critical value, the transition temperature dropped rapidly with decreasing stress (fig. 28).

A suggestion of Fisher's (1954) indicates a reason why such a transverse stress may be needed. The cleavage plane will be threaded by dislocations, and each time the crack crosses a screw dislocation a step in the cleavage plane will be produced ;† this will increase the area of the fracture surface and so of the surface energy which must be supplied. If there are  $\rho$  dislocations per unit area, the surface energy will be increased by an amount  $\frac{1}{2}\rho c^2 b \gamma$  ; adding this term to the right of eqn. (2), we see that the

---

† This is probably the origin of the 'river pattern' markings seen on the cleavage surface (fig. 29, Pl. 17).

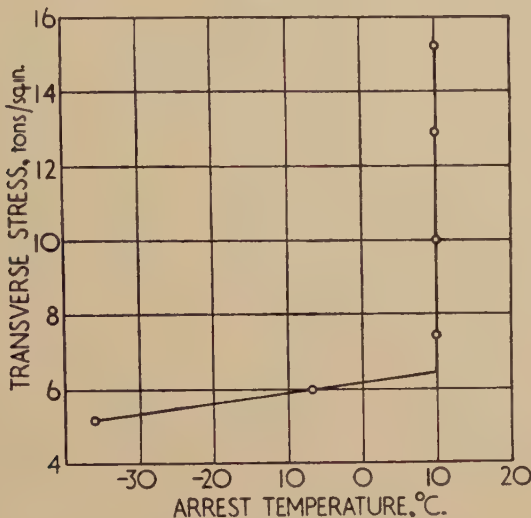


stress needed to spread the crack does not become vanishingly small for large crack, but tends to the limiting value of

$$\sigma = \{4G\rho b\gamma/\pi(1-\nu)\}^{1/2}. \quad (30)$$

With  $\rho \simeq 10^{10} \text{ cm}^{-2}$ , this stress becomes  $6 \times 10^8 \text{ dynes/cm}^2$  or  $4 \text{ tons/in.}^2$ , which is the order of the critical stresses observed by Robertson. If the stress is only slightly above the value (30), then the crack is likely

Fig. 28



Dependence of the arrest temperature of a crack on transverse stress.  
Robertson (1953).

to be held up whenever it meets a density of dislocations greater than the average; for example, when the crack crosses a low angle mosaic boundary. When this happens the delay may give an opportunity for plastic flow to occupy and so to prevent further spread of the crack. This plastic flow can be avoided only if the temperature is sufficiently low; hence a decrease in stress will be accompanied by a fall in the temperature of arrest.

## § 7. DUCTILE FRACTURE

If a ductile specimen is tested in tension three stages in its failure are observed.

(i) The specimen necks. From this point on further deformation is confined to the neighbourhood of the neck.

(ii) A crack is formed in the centre of the neck normal to the stress axis (fig. 30, Pl. 17).

(iii) Final separation occurs by shear at an angle of  $45^\circ$  to the stress axis.

The maximum load occurs when necking first starts ; the corresponding stress is the ' ultimate tensile strength ' of the material. This, however, has no direct relationship with fracture but is the point at which uniform elongation becomes unstable. The strain at which this happens can be determined by a simple geometrical construction from the stress-strain curve as was shown by Considère as early as 1885 ; good accounts of this topic (e.g. Orowan 1949) are available, and it will not be discussed here.

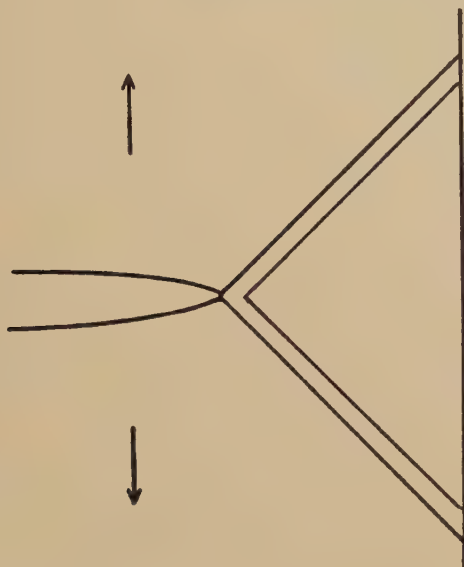
The final stage in failure (iii), is also not a true fracture but is due to localized slip. The explanation of this was first given by Zener (1948) who has also studied a number of other examples of the localization of slip. In general if the strain rate is sufficiently high the deformation will be adiabatic and the resulting rise in temperature produces a softening of the material which acts in opposition to the work hardening. Thus, as Zener and Holloman (1944) have noted, the stress-strain curve has a maximum. For strains greater than that corresponding to the maximum stress, a homogeneous deformation becomes unstable ; the shear strain is localized to the vicinity of a surface which becomes very hot and thereby loses most of its resistance to further deformation. Zener (1948) has demonstrated this in experiments on punching a steel plate (fig. 31, Pl. 17). In 'static' punching deformation occurs generally, but in dynamic punching it is confined to the immediate vicinity of the punching surface. An indication of the high temperatures which are obtained in such dynamic experiments is given by the formation of martensite at the surface.

To see how the cone of a cup and cone fracture may be produced in this way, it must be remembered that, in the final stage of failure, plastic deformation is confined to a small region in the neck : the strain rate here must be large to maintain the average strain rate in the specimen, and so the conditions for adiabatic strain will prevail. When an internal crack has formed plastic strain will be favoured in the two shaded regions in fig. 32 ; under adiabatic straining the plastic strain will be entirely confined to these two regions, and the specimen will separate by slipping apart here.

The formation of the internal crack (stage (ii)) which is the only stage involving a true fracture, has still to be considered. Because the formation of the crack has been preceded by necking, the stress system prevailing is rather complicated. In most experimental investigations the true stress has not been measured but rather the ' conventional stress ' (i.e. the load divided by the original cross sectional area) ; though the latter may be sufficient for technological applications it does not give much information on the fundamental processes involved. One set of experiments in which necking has been fully allowed for and the true stress determined, is due to Petch (1956 a) ; he finds that the maximum stress in the ductile fracture of iron has the same linear dependance on  $d^{-1/2}$ , where  $d$  is the grain diameter, as was found for brittle fracture

(fig. 6); and also the slopes of the lines are the same in both cases. This seems a clear indication that the same basic mechanism must operate, namely that cracks arise from dislocations piled up on the grain boundary. Mott (1953) had previously suggested that this would be the case; in a ductile material, however, the cracks would be stopped by plastic flow from growing to any size, but as the deformation proceeded the number of cracks would increase; ultimately these cracks would link up to form a macroscopic crack.

Fig. 32



Localization of slip near an internal crack. Zener (1948).

Stroh (1957) has suggested that the changes in density and electrical resistance observed by Clareborough *et al.* (1955, 1956) in nickel after heavy cold work can be attributed to microcracks formed in this way. However Seeger (1956) disagrees with this suggestion and maintains that dislocations are sufficient to account for the effects, but to do so he finds it necessary to attribute a rather large electrical resistance to a stacking fault.

The fact that the fracture strength is directly related to the grain size implies that the grain boundaries are the only obstacles strong enough to withstand the large piled up groups needed for fracture. It is noteworthy that single crystals of metals with a face-centred cubic structure do not fracture but separate by drawing down to a double wedge. (Single crystals of other metals do, however, fracture; at present the mechanism by which this occurs has not been established.) In accord with this Stroh (1956) has estimated that one type of obstacle considered important



in work hardening—the sessile dislocation of Lomer (1951) and Cottrell (1952)—is too weak by an order of magnitude to withstand the large piled up group needed for fracture.

Figure 6 shows that the intercept  $\sigma_0$  varies considerably with the conditions under which fracture occurs. (The change shown is probably due not only to the different temperatures, but also to the fact that different amounts of plastic strain preceded fracture in each case.)  $\sigma_0$  has been interpreted in § 4.2 as the stress needed to move a dislocation, and as such it will be affected by all the factors determining the flow stress in a work hardened material†. No theoretical treatment of  $\sigma_0$  in the ductile case has as yet been given, while experimentally only the two values shown in fig. 6 have been determined; very little can be said, therefore, about the variation of  $\sigma_0$ . It may be surmized, however, that most processes effecting the fracture strength (for example the effect of prior plastic deformation which has been studied by Ripling and Baldwin (1951) under the name of rheotropic embrittlement) will do so by producing a change in the value of  $\sigma_0$ .

#### ACKNOWLEDGMENTS

My thanks are specially due to Professor N. F. Mott, F.R.S., for his encouragement in the writing of this article, and for valuable discussions and comments. I also wish to thank Dr. A. R. Entwisle, Dr. E. O. Hall and Dr. P. Pratt for most helpful discussions, and Dr. J. Nutting for permission to reproduce fig. 4, Pl. 14.

### APPENDIX A

#### *The Energy of a Crack*

The change in the elastic energy which occurs when a crack is formed near a piled up group (i.e. along the plane OP of fig. 2) is considered here. Equation (5) for the stress on OP may be written

$$\sigma = A/r^{1/2}, \quad . . . . . (A1)$$

where  $A = L^{1/2}f(\theta) = \frac{3}{2}\sigma_s L^{1/2} \sin \theta \cos \frac{1}{2}\theta, \quad . . . . . (A2)$

when the value of  $f(\theta)$  is inserted. If a crack of length  $r$  is formed along OP, but the stresses are not allowed to relax, (A1) gives the normal stress on the surface of the crack. Now take as origin of coordinates the centre of the crack,  $y$  axis in the plane of the crack, and unit of length, one half the length of the crack. The stress (A1) on the surface of the crack becomes

$$\sigma(y) = A(1-y)^{-1/2}, \quad . . . . . (A3)$$

If the stress (A3) is relaxed the change in the elastic energy (Stroh 1955 a) is

$$W_e = -\pi(1-\nu)G^{-1} \int_0^1 (\tau'^2 + \tau''^2) \mu \, d\mu, \quad . . . . . (A4)$$

† Recent accounts of the theory of work hardening have been given by Friedel (1957) and Seeger *et al.* (1957).



The stationary values of  $W$  occur when

$$c^2 = \{1 - 2(c_1/c_2)^{1/2} \sigma_1 (\sigma_1^2 + \sigma_2^2)^{-1/2} \sin \theta\} c_2 c + c_1 c_2 = 0. \quad (B5)$$

Equation (B5) has either two positive real roots, or no real roots. If there are two real roots, the smaller gives a minimum value of the energy  $W$ , and so a crack of this length will remain in stable equilibrium in the material. If there are no real roots no position of equilibrium of the crack is possible, and so it will grow indefinitely. The critical case occurs when the roots of (B5) are equal; this will happen if

$$2(c_1/c_2)(1 + \sigma_1(\sigma_1^2 + \sigma_2^2)^{-1/2} \sin \theta) = 1,$$

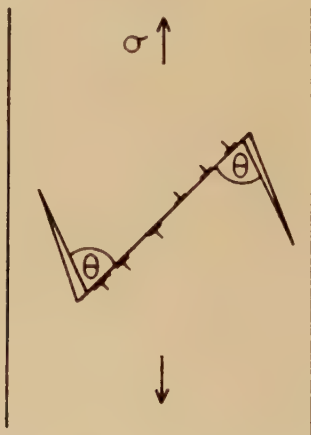
or using (B2) and (B3), if

$$n'b\{(\sigma_1^2 + \sigma_2^2)^{1/2} + \sigma_1 \sin \theta\} = 4\gamma. \quad (B6)$$

The condition for the initiation of the crack, eqn. (9),  $nb\sigma_s = \frac{3}{8}\pi^2\gamma$ , must also be satisfied, where  $n$  is the total number of dislocations in the piled up group and  $\sigma_s$ , the resolved shear stress on the slip plane. Hence the formation of the crack will result in immediate fractions of the material if

$$n'/n = 32\sigma_s/3\pi^2\{(\sigma_1^2 + \sigma_2^2)^{1/2} + \sigma_1 \sin \theta\} \quad (B7)$$

Fig. 33



Slip line and initial crack in a tensile specimen.

Now suppose the applied stress is a uniaxial tension  $\sigma$ . The most favoured slip lines are those at an angle of  $45^\circ$  to the tensile axis, so

$$\sigma_s = \frac{1}{2}\sigma. \quad (B8)$$

The calculations of Stroh (1954) indicate that the crack should form at an angle of about  $\theta = \cos^{-1} \frac{1}{3} = 70.5^\circ$  to the slip plane; in this case we find

$$\sigma_1 = \frac{1}{2}(1 - 4/9\sqrt{2})\sigma = 0.186\sigma, \quad \text{and} \quad \sigma_2 = \frac{7}{18}\sigma. \quad (B9)$$



$$n'/n=0.89; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (B10)$$

In the above treatment we have neglected the stress due to those dislocation of the piled up group which have not entered the crack; since these will tend to spread the crack, consideration of their effect can only reinforce our conclusion that the crack will be unable to find an equilibrium position and so will spread catastrophically.

ALLEN, N. P., HOPKINS, B. E., and MCLENNON, J. E., 1955, *Proc. roy. Soc. A*, **234**, 221.

BAEYERTZ, M., CRAIG, W. F., and BUMPS, E. S., 1949 *Trans. Amer. Inst. min. (metall.) Engrs.* **185**, 481.

BECHTOLD, J. H., 1953, *Trans. Amer. Inst. min. (metall.) Engrs*, **197**, 1469 ; 1954, *Trans. Amer. Soc. Metals*, **46**, 1449 ; 1955, *Acta Met.*, **3**, 249.

BILBY, B. A., and BULLOUGH, R., 1954, *Phil. Mag.*, **45**, 631.

BOODBERG, A., DAVIS, H. E., and PARKER, E. R., 1948, *Weld. J.*, **27**, 186 s.

BUECHE, A. M., and WHITE, A. V., 1956, *J. appl. Phys.*, **27**, 980.

CHURCHMAN, A. T., 1955, *Acta Met.*, 3, 22.

CLARK, D. S., and WOOD, D. S., 1949, *Proc. Amer. Soc. Test. Mater.*, **49**, 717.

CLAREBROUGH, L. M., HARGREAVES, M. W., and WEST, G. W., 1955, *Proc. roy. Soc. A*, **232**, 252; 1956, *Phil. Mag.*, **1**, 528.

CONSIDÈRE, A., 1885, *Ann. des ponts et chaussées*, 6, 9.

COTTRELL, A. H., 1948 *Report of Bristol Conference on the Strength of Solids* (London: Physical Society), p. 30; 1949, *Progress in Metal Physics I*

(London : Physical Society), p. 77 ; 1952 *Phil. Mag.*, **43**, 645 ; 1953 (London : Pergamon Press), p. 77 ; 1952 *Phil. Mag.*, **43**, 645 ; 1953 *Dislocations and Plastic Flow on Crystals* (Oxford : University Press).

COTTRELL, A. H., and BILBY, B. A., 1949, *Proc. phys. Soc. Lond. A*, **62**, 49.

CRACKNELL, A., and PETCH, N. J., 1955, *Acta Met.*, 3, 186.

CRUSSARD, C., BORIONE, R., PLATEAU, J., MORRILLON, Y., and MARATRAY, F.,  
1956 *J. Iron St. Inst.*, **183**, 146.

DERUYTÈRE, A., and GREENOUGH, G. B., 1954, *Phil. Mag.*, **45**, 624 ; 1955, *J. Inst. Met.*, **84**, 337.

ESHELBY, J. D., FRANK, F. C., and NABARRO, F. R. N., 1951, *Phil. Mag.*, **42**, 351.

FELBECK, D. K., and OROWAN, E., 1955, *Weld. J.*, **34**, 570 s.

FISHER, J. C., 1954, Unpublished communication to Birmingham Conference; 1955, *Acta Met.*, **3**, 109.

FRANK, F. C., 1950, *Symposium on the plastic deformation of crystalline solids* (Pittsburgh: Carnegie Institute of Technology and Office of Naval Research), p. 89.

- FRIEDEL, J., 1956, *Les Dislocations* (Paris : Gauthier-Villars) ; 1957, *Proc. roy. Soc. A*. To be published.
- GIBBONS, D. F., and COTTRELL, A. H., 1947, *Nature, Lond.*, **162**, 488.
- GILMAN, J. J., 1956, *J. appl. Phys.*, **27**, 1262.
- GREEN, A. P., and HUNDY, B. B., 1956, *J. Mech. phys. Solids*, **4**, 128.
- GREENWOOD, G. W., and QUARREL, A. G., 1954, *J. Inst. Met.*, **82**, 551.
- GRIFFITH, A. A., 1920, *Phil. Trans. A*, **220**, 163.
- HALL, E. O., 1951, *Proc. phys. Soc. Lond. B*, **64**, 747 ; 1953, *J. Mech. phys. Solids*, **1**, 227.
- HAUSER, F. E., LANDON, P. R., and DORN, J. E., 1956, *Trans. Amer. Inst. min. (metall.) Engrs*, **206**, 589 ; 1957, *Trans. Amer. Soc. Metals* (to be published).
- HESLOP, J., and PETCH, N. J., 1956, *Phil. Mag.*, **1**, 866.
- HILL, R., 1949, *Quart. J. Mech. appl. Math.*, **2**, 40.
- HODGE, J. M., MANNING, R. D., and REICHHOLD, H. M., 1949, *Trans. Amer. Inst. min. (metall.) Engrs*, **185**, 233.
- HOLLOMAN, J. H., 1944, *Trans. Amer. Inst. min (metall.) Engrs*, **158**, 298.
- HUDSON, G., and GREENFIELD, M., 1947, *J. appl. Phys.*, **18**, 405.
- INGLIS, C. E., 1913, *Trans. Instn nav. Archit., Lond.*, **55**, 219.
- JACQUET, P. A., 1954, *Acta Met.*, **2**, 752.
- KENNEDY, H. E., 1945, *Weld. J.*, **10**, 597 s.
- KOCHENDÖRFER, A., 1954, *Arch. Eisenhüttenw.*, **25**, 351.
- LOMER, W. M., 1951, *Phil. Mag.*, **42**, 1327.
- LOVE, A. E. H., 1927, *Theory of Elasticity* (Cambridge : University Press).
- LOW, J. R., 1954, *Trans. Amer. Soc. Metals*, **46 a**, 163.
- LUDWIK, P., 1926, *Z. Metallk.*, **18**, 269.
- LUDWIK, P., and SCHEU, R., 1923, *Stahl u. Eisen, Düsseldorf*, **43**, 999.
- MCADAM, D. J., GEIL, G. W., and MEBS, R. W., 1947, *Trans. Amer. Inst. min. (metall.) Engrs*, **172**, 323.
- MCADAM, D. J., and MEBS, R. W., 1943, *Proc. Amer. Soc. Test. Mater.*, **43**, 661.
- MACGREGOR, C. W., and GROSSMAN, N., 1952, *Weld. J.*, **31**, 20.
- MESNAGER, A., 1906, *Intern. Ass. Testing Materials, Brussels Cong., Rapport non-offic. No. A. 6f.*, p. 1.
- MOTT, N. F., 1948, *Engineering, Lond.*, **165**, 16 ; 1952, *Imperfections in Nearly-perfect Crystals*, ed. W. Shockley (New York : John Wiley), p. 173 ; 1953, *Proc. roy. Soc. A*, **220**, 1 ; 1956, *J. Iron St. Inst.*, **183**, 233.
- MOTT, N. F., and NABARRO, F. R. N., 1948, *Report of Bristol Conference on Strength of Solids* (London : Physical Society), p. 1.
- OROWAN, E., 1933, *Z. Phys.*, **86**, 195 ; 1945, *Trans. Instn Engrs Shipb. Scot.*, **89**, 165 ; 1949, *Rept. Progr. Phys.*, **12**, 185 ; 1955, *Weld. J.*, **34**, 157 s.
- OROWAN, E., NYE, J. F., and CAIRNS, W. J., 1945, *Theoret. Res. Rep. No. 16* 45, Arm. Res. Dep., M.O.S., London.
- PETCH, N. J., 1953, *J. Iron St. Inst.*, **174**, 25 ; 1954, *Progress in Metal Physics V* (London : Pergamon Press), p. 1 ; 1956 a, *Phil. Mag.*, **1**, 186 ; 1956 b, *Ibid.*, **1**, 331.
- PETCH, N. J., and STABLES, P., 1952, *Nature, Lond.*, **169**, 842.
- PRANDTL, L., 1923, *Z. angew. Math. Mech.*, **3**, 401.
- READ, W. T., 1953, *Dislocations in Crystals*, (New York : McGraw Hill).
- RINEBOLT, J. A., and HARRIS, W. J., 1951, *Trans. Amer. Soc. Metals*, **43**, 1175.
- RIPLING, E. J., and BALDWIN, W. M., 1951, *Trans. Amer. Soc. Metals*, **43**, 778.
- ROBERTS, D. K., and WELLS, A. A., 1954, *Engineering, Lond.*, **178**, 820.
- ROBERTSON, T. S., 1951, *Engineering, Lond.*, **172**, 445 ; 1953, *J. Iron St. Inst.*, **175**, 361.
- SACK, R. A., 1946, *Proc. phys. Soc. Lond.*, **58**, 729.

- SCHARDIN, H., and STRUTH, W., 1938, *Glastech. Ber.*, **16**, 219.
- SEEGER, A., 1956, *Canad. J. Phys.*, **34**, 1219.
- SEEGER, A., DIEHL, J., MADER, S., and REBSTOCK, H., 1957, *Phil. Mag.*, **2**, 323.
- SMITH, H. L., and FERGUSON, W. J., 1950, *N.R.L. Progress Report*.
- SMITH, R. L., and RUTHERFORD, J. L., 1957, *Trans. Amer. Inst. min. (metall.) Engrs.*, **209**, 478.
- STROH, A. N., 1954, *Proc. roy. Soc. A*, **223**, 404 ; 1955 a, *Ibid.*, **232**, 548 ; 1955 b, *Phil. Mag.*, **46**, 198 ; 1956, *Ibid.*, **1**, 489 ; 1957, *Ibid.*, **2**, 1.
- TAYLOR, G. I., 1946, *J. Instn civ. Engrs*, **26**, 486.
- TAYLOR, J. W., 1954, *Metallurgia*, **59**, 161.
- TIPPER, C. F., 1945, *Report of Conf. on Brittle Fracture in Steel Plates, Cambridge* (British Iron and Steel Res. Assoc.), p. 24 ; 1957, *J. Iron. St. Inst.*, **185**, 4.
- TIPPER, C. F., and HALL, E. O., 1953, *J. Iron St. Inst.*, **175**, 9.
- TRURY, R., and KRAUS, S., 1936, *Nature, Lond*, **138**, 331 ; 1937, *Ibid.*, **139**, 30.
- ULMO, J., and BASTENAIRE, F., 1953, *Rev. Stat. appl.*, **1**, 45.
- VANDERBECK, R. W., WILDE, H. D., LINDSAY, R. W., DANIEL, C., 1953, *Weld. J.*, **32**, 325 s.
- VON MISES, R., 1928, *Z. angew. Math. Mech.*, **8**, 161.
- WAIN, H. L., and COTTRELL, A. H., 1950, *Proc. phys. Soc. Lond. B*, **63**, 339.
- WAIN, H. L., and HENDERSON, F., 1953, *Proc. phys. Soc. Lond. B*, **66**, 515.
- WITTMAN, F., and STEPANOV, W., 1939, *J. Tech. phys. U.S.S.R.*, **9**, 1070.
- ZENER, C., 1948, *Fracturing of Metals* (Amer. Soc. Metals), p. 3.
- ZENER, C., and HOLLOMAN, J. H., 1944, *J. appl. Phys.*, **15**, 22.
- ZWICKY, F., 1923, *Z. Phys.*, **24**, 131.





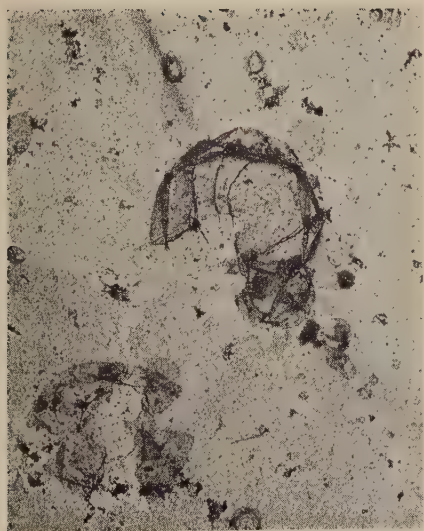
Plate I



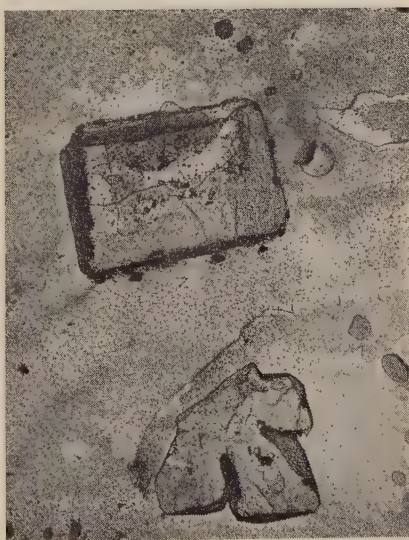
*From Brit. J. appl. Phys.*

Distribution of opposite charges on polythene sheet rubbed on another piece of the same material.

Plate III



(a)



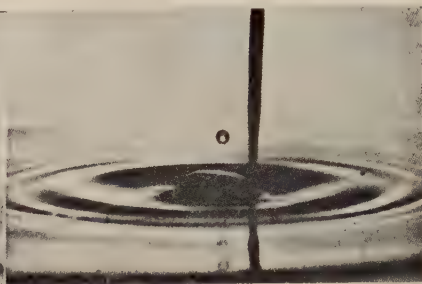
(b)

- (a) Veils from crushed ice (somewhat damaged).  
(b) Veils from sodium chloride crystals.

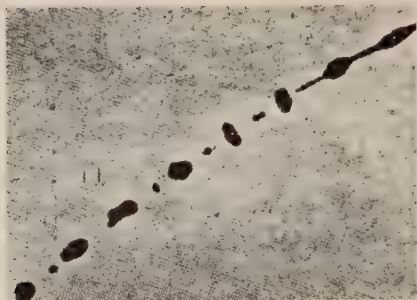
Plate II



*From Trans. Instn chem. Engrs.*  
(a)



*From Trans. Instn chem. Engrs.*  
(b)



*From A University Text-book of Physics (Griffin & Co.)*  
(c)



(d)



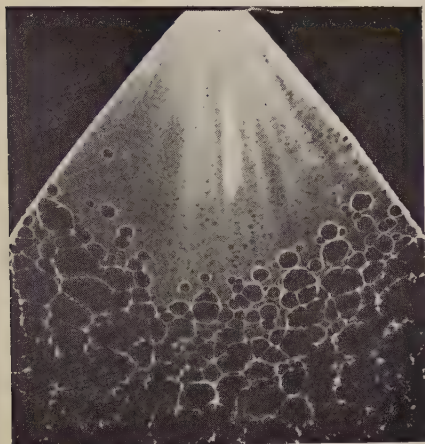
*From Trans. Instn chem. Engrs.*  
(e)



*From Trans. Instn chem. Engrs.*  
(f)

- (a) Blister on bursting bubble.
- (b) Droplet formed from jet following bubble burst.
- (c) Plateau's spherules between droplets formed from jet.
- (d) Shattering of a drop in an air stream.
- (e) Disintegration of filament blown off baffle (low speed).
- (f) Sheet of liquid blown off baffle (high speed).

Plate II



*From Phil. Trans.*

(g)



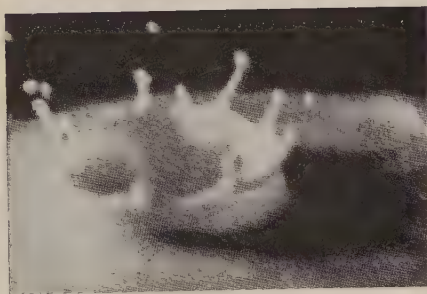
*From Phil. Trans.*

(h)



*From Phil. Trans.*

(j)



*By courtesy of Society for Promoting Christian Knowledge.*

(k)

- (g) Lace-like formation in sheet of liquid from jet.
- (h) Filaments torn from sheet of liquid from jet.
- (j) Direct disintegration of sheet of liquid from jet.
- (k) Droplets detaching from sheet of liquid resulting from splash of drop.

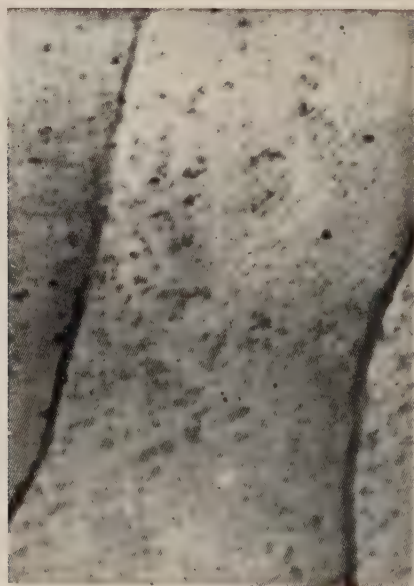


Fig. 3



Piled up groups of dislocations in polycrystalline  $\alpha$ -brass. Jacquet (1954).

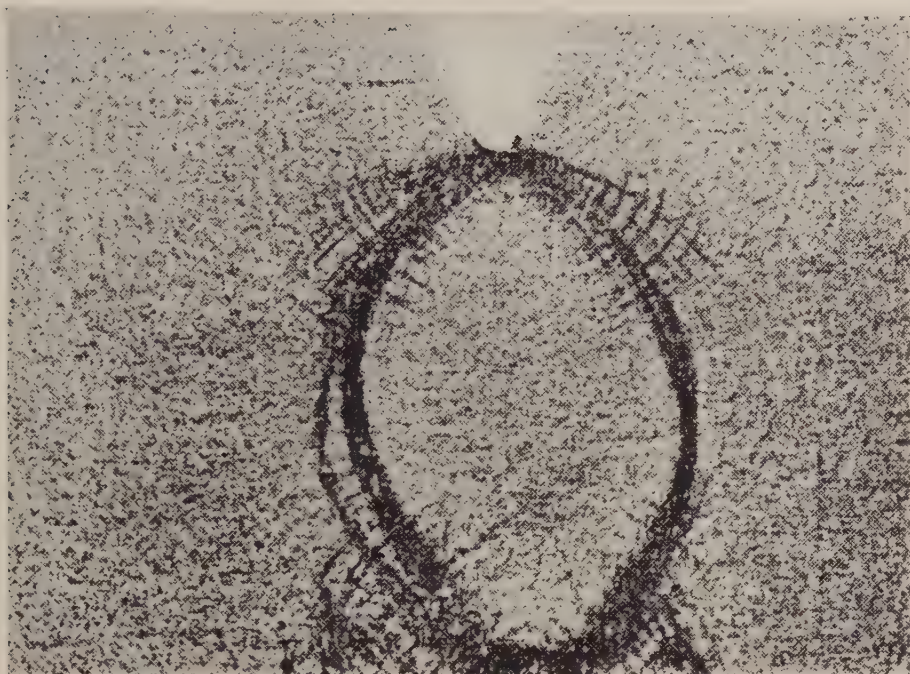
Fig. 4



Slip lines in iron. Direct carbon replica.  $\times 25,000$ . Nutting and Brandon (private communication).



Fig. 22



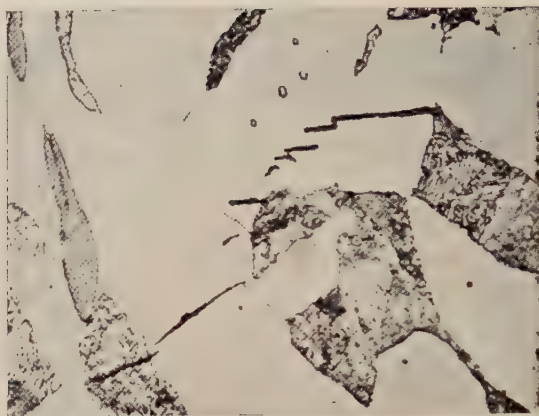
Deformation pattern observed after pure bending. Green and Hundy (1956).

Fig. 23



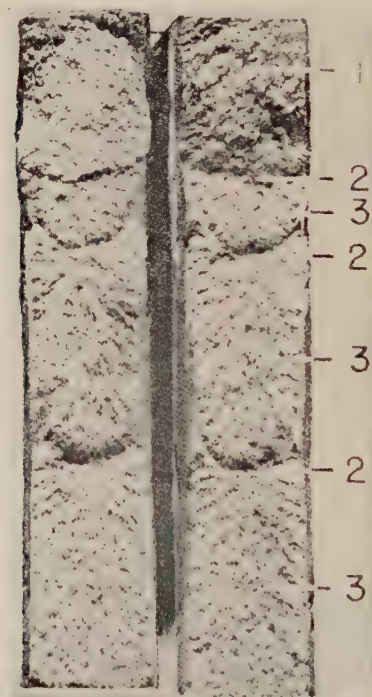
Section of a cleavage plane in an iron crystal  $\times 200$ . Tipper and Hall (1953).

Fig. 25



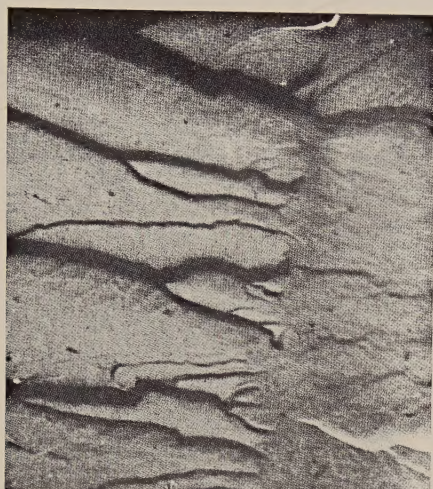
Typical groups of microcracks ahead of the main crack in mild steel plate  $\times 700$ . Tipper (1957).

Fig. 27



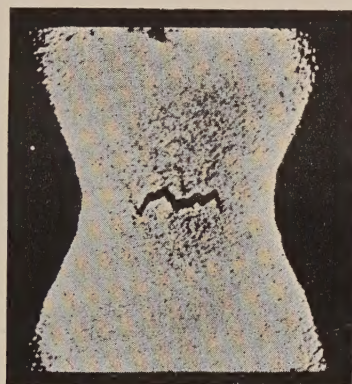
The propagation of a crack in ship-plate showing changes from ductile to brittle behaviour. Felbeck and Orowan (1955).

Fig. 29



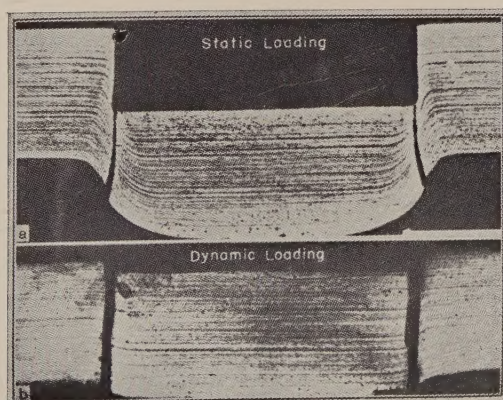
'River pattern' marking on pearlite free steel fractured at  $-196^{\circ}\text{C}$ .  
Crussard *et al.* (1956).

Fig. 30



Section of a specimen showing internal crack formed during a tensile test.  
Ludwik (1926).

Fig. 31



Experiments on punching. Zener (1948).







# Annals of Science

A QUARTERLY REVIEW OF  
THE HISTORY OF SCIENCE  
SINCE THE RENAISSANCE

## EDITORS

D. McKIE, D.Sc., Ph.D.,  
University College, London.

HARCOURT BROWN,  
M.A., Ph.D.,  
Brown University, Providence, R.I.,  
U.S.A.

H. W. ROBINSON,  
Former Librarian,  
Royal Society of London.

N. H. de V. HEATHCOTE,  
B.Sc., Ph.D.,  
University College, London.

ANNUAL SUBSCRIPTION

£3 3s. 0d.

OR

18s. 0d.

PER PART  
POST FREE



## THE MATHEMATICAL WORKS OF JOHN WALLIS, D.D., F.R.S.

by

J. F. SCOTT, Ph.D., B.A.

"His work will be indispensable to those interested in the early history of The Royal Society. I commend to all students of the Seventeenth Century, whether scientific or humane, this learned and lucid book."—Extract from foreword by Prof. E. N. da C. Andrade, D.Sc., Ph.D., F.R.S.

Recommended for publication by University of London

12/6 net

Printed and Published by

**TAYLOR & FRANCIS, LTD.**  
RED LION COURT, FLEET STREET, LONDON, E.C.4.

# Early Scientific Publications



**DIARY OF ROBERT HOOKE, M.A., M.D., F.R.S.**

1672-1680

Edited by **H. W. ROBINSON** and **W. ADAMS**  
Recommended for publication by the Royal Society,  
London

**25/-**

net

"This vivid record of the scientific, artistic and social activities of a remarkable man during remarkable years has too long remained in obscurity."—Extract from foreword by Sir Frederick Gowland Hopkins, O.M., President of the Royal Society.

**MATHEMATICAL WORK OF JOHN WALLIS, D.D., F.R.S.**

By **J. F. SCOTT, Ph.D., B.A.**

**12/6**

net

"His work will be indispensable to those interested in the early history of The Royal Society. I commend to all students of the Seventeenth Century, whether scientific or humane, this learned and lucid book."—Extract from foreword by Prof. E. N. da C. Andrade, D.Sc., Ph.D., F.R.S.  
Recommended for publication by University of London

**CORRESPONDENCE AND PAPERS OF EDMOND HALLEY**

**21/-**

net

Arranged and Edited by **EUGENE FAIRFIELD MACPIKE**

First published on behalf of The History of Science Society by Oxford University Press. Now re-issued by Taylor & Francis, Ltd.

**MEMOIRS OF SIR ISAAC NEWTON'S LIFE**

**5/-**

net

By **WILLIAM STUKELEY, M.D., F.R.S., 1752**

From an Original Manuscript

Now in the possession of the Royal Society, London

**HEVELIUS, FLAMSTEED AND HALLEY**

**12/6**

net

Three Contemporary Astronomers and their Mutual Relations  
By **EUGENE FAIRFIELD MACPIKE**

Published by arrangement with The History of Science Society

Established  
over 150 years

**TAYLOR & FRANCIS, LTD.**  
RED LION COURT, FLEET ST., LONDON E.C.  
PRINTERS & PUBLISHERS OF SCIENTIFIC BOOKS